

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OH nucleic - nucleic search, using sw model

Run on: September 17, 2002, 05:04:18, Search time 258.22 seconds

(without alignments)  
4179,050 Million cell updates/sec

Title: US-09-863-823-5

Sequence: 1 atggcattgaagagagcagcgtt.....acagtgaacacgctctatga 789

Scoring table: IDENTITY\_NMC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

EST:  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpj:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: qb\_estli:\*  
10: qb\_estli2:\*  
11: qb\_hic:\*  
12: qb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vitc:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	671.4	85.1	822	10	BT64721 603051058
2	504.6	64.0	1367	11	AK009118 Mus muscu
3	487.2	61.7	1317	11	AK008028 Mus muscu
4	435.4	55.2	797	10	BG970132 B970132
5	425.6	53.9	646	10	AA986919 uc80802.y
6	415.4	52.7	964	10	PF785067 BF785067
7	402.6	51.0	538	10	BM445724 1115A1C
8	382.2	48.4	936	10	BF788117 602113341
9	381.8	48.4	543	9	AA500290 v197g08.r
10	336.2	42.6	657	9	A1747948 u104c10.y
11	306.6	38.9	445	9	A1956638 u179f12.y
12	296.6	37.5	487	9	A1615364 v197g08.y
13	241.2	30.6	514	9	AW211197 u125h03.y
14	215.4	27.3	437	9	BB845854 BB845854
15	189.8	24.1	515	12	A2864346 CM0173F20
16	160	20.3	504	9	AW235098 x118e10.x
17	156.8	19.9	175	9	AM844313 RC2-CN005

18	131.8	16.7	535	9	AW147007
19	108	13.7	504	10	W64090
20	106.4	13.5	320	9	A1425843
21	90.6	11.5	125	10	DM45552
22	61.2	7.8	548	10	DM026157
23	51.4	6.5	572	9	A1790737
24	43.8	5.6	590	10	BM025920
25	43.8	5.5	151	7	AA463845
26	44.4	5.5	522	9	AW107893
27	40.6	5.1	433	9	AA772324
28	40.6	5.1	780	12	CNS04142
29	40.2	5.1	393	10	W94303
30	40.2	5.1	411	12	A2979833
31	39.8	5.0	435	9	A1672771
32	39.2	5.0	433	9	A1032799
33	39.2	5.0	513	9	A1160022
34	39.2	5.0	523	10	BF063348
35	39.2	5.0	637	10	BE350315
36	39.2	5.0	670	10	BM171749
37	39.2	5.0	816	9	AU117395
38	39.2	4.9	252	10	BF785067
39	39.2	4.9	474	9	BB818563
40	39.2	4.9	510	14	FK002197
41	39.2	4.9	642	9	BB652201
42	39.2	4.9	665	9	BB614047
43	39.2	4.9	676	9	BB465902
44	39.2	4.9	940	11	AK020591
45	38.2	4.8	478	9	AA844204

## ALIGNMENTS

RESULT 1  
BT64721 822 bp. mRNA linear. EST 25 SEP 2001  
LOCUS 603051058F1 NIH\_MGC\_116 Homo sapiens CDNA clone IMAGE:5190876 5,  
DEFINITION mRNA sequence.  
ACCESSION BT64721 GI:15756299  
VERSION BT64721.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 822)  
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINT at:  
http://image.llnl.gov  
Plate: L1AM11477 row: d column: 15  
High quality sequence slp: 821.  
Location/Qualifiers  
1..822  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGF:190878"  
/clone\_id="NIH\_MGC\_116"  
/lab\_host="DH104"  
/note="Organ: pooled colon, kidney, stomach; Vector:  
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
source anonymous pool of 3 colons, age 26 yo male, 49 yo  
female, 71 yo male colon, 46 yo male kidney, and pool of 2  
stomachs, 62 yo male and 70 yo female. Library is  
oligo-dT primed and directionally cloned (EcoRV site is























```

UY 103 gtttaaacctgagactatatactctgatactacactggtcccaagcatctctatatt 102
DB 167 ccttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 236
UY 163 tcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 222
DB 222 gttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 286
UY 223 gattttggaattctgaaatgaaatcttcttcttcttcttcttcttcttcttctt 282
DB 287 gatttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 346
UY 283 aattgaaacagagatcagccttcttcttcttcttcttcttcttcttcttcttctt 340
DB 447 atttgaattatgagcttcttcttcttcttcttcttcttcttcttcttcttctt 405
UY 341 cagtgatgctgagatgattcttcttcttcttcttcttcttcttcttcttcttctt 371
DB 407 cagtactgctgcaaccttcttcttcttcttcttcttcttcttcttcttcttctt 437

```

## RESULT 15

AZ864346 515 bp DNA linear GSE 21-PPH-2001

DEFINITION 2M0173F20R Mouse 10kb plasmid 000201M library Mus musculus genomic

ACCESSION AZ864346

VERSION AZ864346.1

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 515)

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, B., Longacre, S., Mahon, M., Meenen, R., Pedersen, T., Reilly, M., Rose, M., Rose, K., Stokes, K., Tinney, A., Von Niederhausen, A., and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL

COMMENT

Unpublished (2000)

Contact: Robert H. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000

Plate: 0173

Seq Error: 0.00

Seq primer: CACACAGCAACACCTATGACC

Class: Plasmid ends

High quality sequence stop: 515.

Location/Qualifiers

## FEATURES

1..515

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="000201M0173F20"

/clone\_lib="Mouse 10kb plasmid 000201M library"

/sox="Male"

/lab\_host="E. coli strain XL10-Gold, T1-resistant, F"

/note="Vector: pW042hy; purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/faq/faq.html). The DNA was hydrolytically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pW042 (941473311494:AT123072.1) a copy number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert 3'UTRs and purified. The sheared, adaptor mouse DNA was annealed to chemically competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 138 a 117 c 105 g 155 t

ORIGIN

Query Match 24.1% Score 189.8; DB 12; Length 515;

Best Local Similarity 81.0% Pred. No. 3.2e-43;

Matches 221; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

UY 378 tggaaacagacttccaaacagllgagagagaglaattgaaatgattgaaatgaa 437

DB 515 TCGCAATGCTTCCAAACCGTGGCAAAACGCAATGTAAGTTGCTTGGCAACGTGAA 456

UY 438 agtcaacacacagcttcaatgagatgaglaaaacagatgctctgatttaagaa 497

DB 455 ATGTAATGCTTCCAAACCGTGGCAAAACGCAATGTAAGTTGCTTGGCAACGTGAA 496

UY 498 aagcagtcacaaatccaaacagagacagagagagagagagagagagagagag 557

DB 495 AGTGGCTTAAATGCAACGCAATGTAAGTTGCTTGGCAACGTGAA 436

UY 558 gaagcctaaacag 617

DB 436 GAAATGCTTCCAAACCGTGGCAAAACGCAATGTAAGTTGCTTGGCAACGTGAA 276

UY 618 ggaatttcaatgagatgagagagagagagagagagagagagagagagagag 650

DB 275 GCACTTCCACTTCTGTTAAAGCTACTCTCT 243

Search completed: September 17, 2002, 09:04:22

Job time: 2575 sec



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: September 17, 2002, 08:21:27 : Search time 2548.21 Seconds

(without alignments)  
4051.931 Million cell updates/sec

Title: US-09-863-823-1

Sequence: 1 atgcaatggaagattct.....acatgaacagctcatga 765

Scoring table: Identity\_Nuc  
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

EST:\*  
1: em\_estba:\*  
2: em\_estbm:\*  
3: em\_estlin:\*  
4: em\_estlov:\*  
5: em\_estlov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estc2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vtl:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	655.8	85.7	822	10	BI764721
2	504.6	66.0	1367	11	AK009118
3	487.2	63.7	1317	11	AK008028
4	455.4	56.9	797	10	BG370132
5	425.6	55.6	646	9	AA986919
6	415.8	54.4	964	10	BF785067
7	402.6	52.6	538	10	BM45724
8	382.2	50.0	936	10	BF788117
9	381.8	49.9	543	9	AA500290
10	336.2	43.9	657	9	AI747948
11	306.6	40.1	445	9	AI956638
12	295.6	38.6	487	9	AI615364
13	241.2	31.5	514	9	AM211197
14	215.4	28.2	437	9	BB845854
15	189.8	24.8	515	12	AZ864346
16	160	20.9	504	9	AM235098
17	156.8	20.5	175	9	AM844313

18	131.8	17.2	535	9	AM107007
19	108	14.1	504	10	W82090
20	106.4	13.9	320	9	AI425843
21	89.8	11.7	125	10	BM445552
22	61.2	8.0	545	10	BM026157
23	52	6.8	572	9	AI790737
24	51.4	6.7	590	10	BM025920
25	43.8	5.7	151	9	AA489845
26	43.4	5.7	524	9	AM107893
27	40.6	5.3	480	9	AA772324
28	40.6	5.3	780	12	CNS04142
29	40.2	5.3	394	10	W84444
30	40.2	5.3	411	12	AC979833
31	39.8	5.2	435	9	AI672771
32	39.2	5.1	433	9	AI032799
33	39.2	5.1	513	9	AI180042
34	39.2	5.1	523	10	BF063348
35	39.2	5.1	637	10	BF350315
36	39.2	5.1	670	10	BM171749
37	39.2	5.1	816	9	AM117395
38	39	5.1	252	10	BF785567
39	39	5.1	474	9	BB818563
40	39	5.1	610	12	FR0021997
41	39	5.1	642	9	BB522301
42	39	5.1	665	9	BB614047
43	39	5.1	676	9	BB465902
44	39	5.1	940	11	AK020591
45	38.2	5.0	478	9	AA042404

#### ALIGNMENTS

RESULT 1	BI764721	822 bp	mRNA	linear	EST 26-Sep-2001
DEFINITION	603051058F1 NIH_MGC_116 Homo sapiens CDNA clone IMAGE:5190878 5'				
ACCESSION	BI764721				
VERSION	BI764721.1	GI:15756299			
KEYWORDS	EST.				
ORGANISM	Homo sapiens				
REFERENCE	NIH-MGC h19p17/mgc not nih gov/				
AUTHORS	1 (bases 1 to 822)				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.jnl.gov Plate: LNA11477 row: d column: 15 High quality sequence stop: 821. Location/Qualifiers 1. 822				

#### FEATURES

source	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:5190878"
	/clone_11b="NIH_MGC_116"
	/lab_host="DH10B"
	/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-Sp6K16, site: 1. NciI, site: 2. EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney; and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is





[illegible][illegible]



1003

1

1



Tel: 780 492 0169  
 Fax: 780 492 4265  
 Email: smoores@us.nslb.ca  
 Insert Length: 538 Std Error: 0.00  
 POLY-A-No.

## FEATURES

## SOURCE

Location/Qualifiers

1..538  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="Bos taurus ileum #1 library"  
 /tissue\_type="Smooth muscle"  
 /cell\_type="Simple columnar epithelial"  
 /dev\_stage="Young adult"  
 /lab\_host="X11-BiuekR strain"  
 /note="Organ: Intestine/Ileum; Vector: Uni-22APXR; Site\_1:  
 EcorI; Site\_2: Xho I"

BASE COUNT 165 a 118 c 121 g 134 t  
 ORIGIN

## Query Match

Best Local Similarity 85.7%; Score 402.6; PB 10; Length 538;  
 Matches 457; Conservative 2; Mismatches 71; Indels 3; Gaps 1;

OY 42 gccacgtgagatgacaggtctgtttaaactggaatgataaacctgagactatctt 101  
 DB 5 gccacgtgagatgacaggtctgtttaaactggaatgataaacctgagactatctt 64  
 OY 102 gatatctacactgctgctcccaacatctgatatgctgcttcaaacacacacagaga 161  
 DB 65 ggacactgagctgctgctcccaacatctgatatgctgcttcaaacacacacagaga 124  
 OY 162 ggaagaaatctgtctgtgaacgagagagagagagagagagagagagagagag 221  
 DB 125 tgaagaaatctgtctgtgaacgagagagagagagagagagagagagagagag 184  
 OY 222 caattcagctctgtctgtctctccatcagtgaaatgagacagagagagagagag 281  
 DB 185 caattcagctctgtctgtctctccatcagtgaaatgagacagagagagagagag 244  
 OY 282 ctgacagctgctgctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 341  
 DB 245 ctgacagctgctgctgtctgtctgtctgtctgtctgtctgtctgtctgtctgt 304  
 OY 342 tctctctaaagtgaagacagcttcaaacagtgagagagagagagagagagagag 401  
 DB 305 tctctctaaagtgaagacagcttcaaacagtgagagagagagagagagagagag 364  
 OY 402 ttgcacatgaaagcaaccccccagctcgaatgagatgagatgagatgagatgag 461  
 DB 365 ttgcacatgaaagcaaccccccagctcgaatgagatgagatgagatgagatgag 424  
 OY 462 cgatttgaagaaagccgctcaccacacacacacacacacacacacacacacac 521  
 DB 425 gaatttgaagaaagccgctcaccacacacacacacacacacacacacacacac 481  
 OY 522 caccacagctcgaagagccttgcacagagagagagagagagagagagagagag 574  
 DB 482 caccacagctcgaagagccttgcacagagagagagagagagagagagagagag 544

## RESULT

BF788117

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

57 mRNA sequence.  
 BF788117.1 GI:12093166  
 EST.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 936)  
 AUTHORS NIH-MGC  
 TITLE NIH-MGC  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: rgap@remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Lucyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
 http://image.llnl.gov  
 Plate: L149857 row: b column: 11  
 High quality sequence start: 5  
 High quality sequence stop: 713.  
 Location/Qualifiers

## FEATURES

## SOURCE

1..936

/organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_lib="IMAGE:4241194"  
 /clone\_lib="NCI-CCAP Kid14"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Kidney; Vector: pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.75 kb. Constructed by Life  
 Technologies. Note: this is a NCI-CCAP library. 1"

## BASE COUNT

270 a 213 c 231 g 220 t

## Query Match

Best Local Similarity 80.8%; Score 387.2; DB 10; Length 936;  
 Matches 492; Conservative 2; Mismatches 110; Indels 5; Gaps 4;

OY 19 ctctcttaagaaattatcttctccacagtgagatgacagctctttaaactgaa 78  
 DB 31 ctctcttaagaaattatcttctccacagtgagatgacagctctttaaactgaa 88  
 OY 79 gtaaaacatgagacatatactctgatactacacacacacacacacacacacacac 138  
 DB 89 gtaaaacatgagacatatactctgatactacacacacacacacacacacacac 148  
 OY 139 gctgttcaaaac 198  
 DB 149 gctgttcaaaac 208  
 OY 199 gatttgaatctgaaac 258  
 DB 209 gatttgaatctgaaac 267  
 OY 259 aatgacaaaggaatcgaacttaactcgaagctgagagagagagagagagagag 318  
 DB 268 aatgacaaaggaatcgaacttaactcgaagctgagagagagagagagagagag 327  
 OY 319 gctgttcaaaac 378  
 DB 328 gctgttcaaaac 387  
 OY 379 gaagcagaaatgag 438  
 DB 388 gaagcagaaatgag 445  
 OY 439 tggatcaaaac 498  
 DB 446 tggatcaaaac 505  
 OY 499 aatgacaaaggaatcgaacttaactcgaagctgagagagagagagagagagag 558  
 DB 506 aatgacaaaggaatcgaacttaactcgaagctgagagagagagagagagagag 565  
 OY 559 tgaatcgaag 618











---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 17, 2002, 13:09:58, Search time 10.65 seconds  
(without alignments) 428.063 Million cell updates/sec

Title: US-09-863-823-2

Perfect score: 1295  
Sequence: 1 M06GRFLVLVILFPEPMTS.....KKKIMKLCMKDKDPHSETAL 254

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: 1: SPREMBL\_19:\*

2: SP\_Archaea:\*

3: SP\_Fungi:\*

4: SP\_Human:\*

5: SP\_Invertebrate:\*

6: SP\_Mammal:\*

7: SP\_Mhc:\*

8: SP\_Organelle:\*

9: SP\_Phage:\*

10: SP\_Plant:\*

11: SP\_Protoct:\*

12: SP\_Virus:\*

13: SP\_Vertebrate:\*

14: SP\_Unclassified:\*

15: SP\_RVirus:\*

16: SP\_Bacteria:\*

17: SP\_Archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	74.6	261	11	Q9D7L8
2	915	70.7	253	11	Q9D8H2
3	175	13.5	604	4	Q96CJ3
4	173.5	13.4	2762	5	P91255
5	172	13.3	2013	11	Q9ERC8
6	171.5	13.2	946	13	007153
7	169.5	13.1	359	5	Q9V6C2
8	169.5	13.1	431	4	Q9N7Z4
9	166	12.8	605	11	Q921P2
10	161	12.4	1011	5	Q24273
11	161	12.4	1079	5	Q9VNP2
12	159	12.3	597	6	Q95LH0
13	158	12.2	1463	11	055124
14	157.5	12.2	1415	5	Q94155
15	157	12.1	467	5	Q9VP08
16	156	12.0	1173	5	P91456

17	155	12.0	595	4	Q96PQ1	Q96PQ1 homo sapien
18	154.5	11.9	1513	13	Q90270	Q90270 brachydanio
19	154	11.9	509	11	Q9E0Y5	Q9E0Y5 mus musculu
20	154	11.9	509	11	Q91YK7	Q91YK7 mus musculu
21	154	11.9	603	5	Q9NKF5	Q9NKF5 drosophila
22	154	11.9	868	11	Q92838	Q92838 ratius norv
23	153.5	11.9	1344	11	Q9Z214	Q9Z214 mus musculu
24	153	11.8	467	4	Q9Y2B6	Q9Y2B6 homo sapien
25	152.5	11.8	850	4	Q9ULF7	Q9ULF7 homo sapien
26	152.5	11.8	5196	5	Q76518	Q76518 caenorhabdi
27	152	11.7	869	4	Q15146	Q15146 homo sapien
28	144	11.5	950	11	Q91095	Q91095 mus musculu
29	149	11.5	868	11	Q91096	Q91096 mus musculu
30	149	11.5	2037	5	Q9V1S8	Q9V1S8 drosophila
31	148.5	11.5	4370	4	Q9H3V5	Q9H3V5 homo sapien
32	147.5	11.4	436	5	Q94224	Q94224 caenorhabdi
33	147	11.4	1445	11	Q63155	Q63155 ratius norv
34	146	11.3	528	5	P91670	P91670 drosophila
35	146	11.3	545	5	Q9VCT4	Q9VCT4 drosophila
36	146	11.3	7962	4	Q10465	Q10465 homo sapien
37	145.5	11.2	773	4	Q9NSW7	Q9NSW7 homo sapien
38	145.5	11.2	1215	5	Q9V787	Q9V787 drosophila
39	145	11.2	951	5	Q19128	Q19128 caenorhabdi
40	144.5	11.2	334	7	002870	002870 gallus gall
41	144.5	11.2	350	7	002869	002869 gallus gall
42	144.5	11.2	693	4	Q9UP01	Q9UP01 homo sapien
43	144.5	11.2	1482	5	Q9V4Y6	Q9V4Y6 drosophila
44	144.5	11.2	6620	4	Q96AA2	Q96AA2 homo sapien
45	144	11.1	454	11	Q91W54	Q91W54 mus musculu

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	261 AA.
Q9D7L8	Q9D7L8			
AC	Q9D7L8:			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
CE	2010002A20R1K PROTEIN.			
GN	2010002A20R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Kentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TOUGH;			
PX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shindagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,			
RA	Saito T., Okazaki Y., Gotohori T., Hono H., Kasuyawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schmitt L.M., Staubli F., Suzuki K., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustingich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez T., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Tycicka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,			
RA	Hayashizaki Y.;			
FT	*Functional annotation of a full-length mouse cDNA collection.*;			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK009118; BAB26083.1;			
DR	EMBL; M01:191351; 2010002A20R1K.			

```

DR InterPro: IPR003599; 19.
DR InterPro: IPR003598; 19_c2.
DR InterPro: IPR003600; 19_MHC.
DR InterPro: IPR003606; 19_MHC.
DR Pfam: PF00047; 19; 2.
DR SMART: SM00409; 19; 2.
DR SMART: SM00408; 19c2; 1.
DR SMART: SM00410; 19c2; 1.
KW Immunoglobulin domain.
SC SEQUENCE: 261 AA; 29050 MM; 1FFIC02D27E83CD CRC64;

Query Match
Host Local Similarity: 75.8%, Score 966; DB 11; Length 261;
Matches 188; Conservative 25; Mismatches 35; Indels 0; Gaps 0,

UY 7 LLLVILFLPREMTSSVLYNCKTENYILDPTTGSQASLICAVONHTREELWYREGRY 66
DB 14 LLLVILSLPQGRSSVLYNCKTENYILDPTGQVQASLECAVONHTREELWYREGIV 73
UY 67 DLKSKKINSSVCVSSISFNNGISFTCLGRQGVSVSVLVNTPPLISGNGFQVE 126
DB 74 DLKSKKINISSVCVSPINSDNGVRFCKLRQGVSVVYLVNTPPLISGNGFQVE 133
UY 127 EDSNVKLVGNVAVANPQAVMMYKNSSLDLKSRHQIQGTSFQSLSTVKEKPNQYTS 186
DB 134 ENSDVSLVCNVASNPQAVMMYKNSALVLEKGRHQIQGTRESFQSLSTVKKKSDNGYTS 193
UY 187 CLASSSLKTFPSLDFHLYVNDKTVGPPIPIIAACVVFILFLCPLGLIARRKKIMLKCKDK 246
DB 194 CLASSSLKMTDFHLYVNDKTVGPPIPIIAACVVFILFLCPLGLIARRKKIMLKCKKN 253
UY 247 DPHSETAL 254
DB 254 DPHSETAL 261

RESULT 2
Q99BH2 PRELIMINARY; PRT; 253 AA.
ID Q99BH2;
DT 01-JUN-2001 (TREMblrel; 17, Created)
DT 01-JUN-2001 (TREMblrel; 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel; 19, Last annotation update)
DB 201002A20R1K PROBLIN.
GN 201002A20R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RA Kawai J., Shigaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa K., Iizawa A., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Akasaki Y., Asajohori T., Imano H., Masuoka T., Saito K.,
RA Kiyama K., Matsuda N.A., Asahara M., Matsuda S., Goshima T.,
RA Fuchisawa M., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuchel P., Lewis S., Matsuo Y., Nakai T., Fesole G., Quackenbush J.,
RA Schramm L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okita T., Furuno M., Aono H., Haidarrelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.D., Bult C., Fletcher C., Fujita M., Galardi M.,
RA Gaston-Jones S., Hill D., Holman M., Hume D.A., Kaniya K., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Scheinbach G., Seya T., Shibata Y., Sakamoto N.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitlaker C., Wilmink L.,
RA Wyszynski B., Yoshida K., Hasegawa T., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.
RT *functional annotation of a full-length mouse cDNA collection.*;

```

```

BL Nature 409:685-690(2001).
DR EMBL: AK008028; BAB25419.1;
DR MGI: MGI-1911851; 2010002A20P11.
DR InterPro: IPR003599; 19.
DR InterPro: IPR003598; 19_c2.
DR InterPro: IPR003606; 19_MHC.
DR Pfam: PF00047; 19; 2.
DR SMART: SM00409; 19; 2.
DR SMART: SM00408; 19c2; 1.
KW Immunoglobulin domain.
SC SEQUENCE: 253 AA; 28372 MM; HCA6FC9C8D716761 CRC64;

Query Match
Host Local Similarity: 76.1%, Score 915; DB 11; Length 253;
Matches 178; Conservative 24; Mismatches 32; Indels 0; Gaps 0,

UY 7 LLLVILFLPREMTSSVLYNCKTENYILDPTTGSQASLICAVONHTREELWYREGRY 66
DB 14 LLLVILSLPQGRSSVLYNCKTENYILDPTGQVQASLECAVONHTREELWYREGIV 73
UY 67 DLKSKKINSSVCVSSISFNNGISFTCLGRQGVSVSVLVNTPPLISGNGFQVE 126
DB 74 DLKSKKINISSVCVSPINSDNGVRFCKLRQGVSVVYLVNTPPLISGNGFQVE 133
UY 127 EDSNVKLVGNVAVANPQAVMMYKNSSLDLKSRHQIQGTSFQSLSTVKEKPNQYTS 186
DB 134 ENSDVSLVCNVASNPQAVMMYKNSALVLEKGRHQIQGTRESFQSLSTVKKKSDNGYTS 193
UY 187 CLASSSLKTFPSLDFHLYVNDKTVGPPIPIIAACVVFILFLCPLGLIARRKKIMLKCKDK 240
DB 194 CLASSSLKMTDFHLYVNDKTVGPPIPIIAACVVFILFLCPLGLIARRKKIMLKCKKN 247

RESULT 3
Q96CJ3 PRELIMINARY; PRT; 604 AA.
ID Q96CJ3;
AC Q96CJ3;
DT 01-DEC-2001 (TREMblrel; 19, Created)
DT 01-DEC-2001 (TREMblrel; 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel; 19, Last annotation update)
DB 201002A20R1K PROBLIN.
GN 201002A20R1K.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY; AND RENAL CELL ADENOCARCINOMA;
RA Strausberg R.;
RA Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: BC014205; AAH14205.1;
FT NON-TER
SC SEQUENCE: 604 AA; 65271 MM; RHP791A4034P546 CRC64;

Query Match
Host Local Similarity: 13.5%, Score 175; DB 4; Length 604;
Matches 60; Conservative 31; Mismatches 88; Indels 32; Gaps 8;

UY 41 QASLILPQGRSSVLYNCKTENYILDPTTGSQASLICAVONHTREELWYREGRY 82
DB 70 QVTLICASGUPPLSTLWMTSHKINISSEKASWITPEKVEILLDGMVYKSHAVSSTLIK 129
UY 83 SISENNGISFTCLGRQGVSVSVLVNTPPLISGNGFQVEEDSNVAVLVNPK 138
DB 130 SIOYTAG-ELYCTASNTIGD---SOSMYLEVYAPKIQDPVAVVMEGNQVITEVF 185
UY 139 ANPQAVMMYKNSSLDLKSRHQIQGTSFQSLSTVKEKPNQYTSCTANSTKTES 197
DB 186 AYPALISLWPKQGLPSSNTSNITVNPSSASTLEVLPUSBNPGNTNIAVKKIQGS 245
UY 198 LDFHLVCKDKTVGPPIPIIAACVVFILFLCPLGLIARRKKIMLKCKKN 224

```

Db 246 LEF1VQAD-TPSSPSIDOVEFYSSTAOYQF 275

RESULT 4  
P91255 PRELIMINARY: PRT: 2762 AA.

AC P91255;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHEICAL 306.5 KDA PROTEIN.  
GN F12F3.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=90069613; PubMed=9851916;  
RA None;  
RT \*Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.\*;  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Fulton B., Wohlmann P.;  
RT \*The sequence of C. elegans cosmid F12F3.\*;  
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RT \*Direct Submission.\*;  
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL: F80022; AAC25846.2;  
KW Hypothetical protein.  
SQ SEQUENCE 2762 AA; 306525 MW; C964CCEB6D322084 CRC64;

Query Match 13.4%; Score 173.5; DB 5; Length 2762;  
Best Local Similarity 24.1%; Pred. No. 4.6e-06;  
Matches 60; Conservative 16; Mismatches 100; Indels 53; Gaps 9;

QY 4 GRPLLVILFLPREMTSSVLTNGKTYILDTTP-----GS 40  
DB 1562 GFNCLIENLGGASASCCVITFNKPA--LQSTPDHSLERNLPTLQKALNESAQAGQ 1619  
QY 41 QASLIQAVONHILHEELLMTYKEE-----GKVLKSGNKNLSSVCSSISENLGISFI 94  
DB 1620 QIMLTRISSRS-ESTVAAFKDDERFESAGRTFISDDKSNKLYV--HAVQSDPTG KYP 1676  
QY 95 C---RIGRQD---VSASVLTNTFPPLLSGNDFQVTEEGSNVAVCNVANKVQVMMW 147  
DB 1677 CAVTNKRYGAEECNVAVEDVTKFTAPSPSATLSDSTALIGHNITLDEKVGSPAPEVSW 1736  
QY 148 YKNSSLDLEKSRHQIQCTSEFSQSLSTVKEKPDNGTSCIAKS-----SLKT 195  
DB 1737 TKDGERISTTRIRIQTODENGCKLISKAESDDMGVVCATSVAGVDSSTSSMYIAKT 1796  
QY 196 ESLDFHLIV 204  
DB 1797 TGTDSHLVI 1805

RESULT 5  
ID Q9ERC8 PRELIMINARY: PRT: 2013 AA.  
AC Q9ERC8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE DOWN SYNDROME CELL ADHESION MOLECULE.  
GN DSCAM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C57BL/6; TISSUE-BRAIN;  
RA Barlow G.M., Lyons G.E., Korenberg J.R.;  
RT \*Down syndrome cell adhesion molecule, a highly conserved mouse  
homolog, dscam, is expressed in the neuronal and adult mouse brain.\*;  
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ICR OUT-BRED;  
RX MEDLINE=21134302; PubMed=11237714;  
RA Agarwala K.L., Ganesh S., Amano K., Suzuki T., Yamakawa K.;  
RT \*DSCAM, a highly conserved gene in mammals, expressed in  
differentiating mouse brain.\*;  
RL Biochem. Biophys. Res. Commun. 281:697-705(2001).  
DR EMBL: AF315558; AAC28796.1;  
DR EMBL: AY005483; AAF99440.1;  
DR MGI: 1196281; Dscam.  
DR InterPro: IPR003961; FN\_III.  
DR InterPro: IPR003599; I9.  
DR InterPro: IPR003598; I9\_C2.  
DR InterPro: IPR003600; I9\_Like.  
DR InterPro: IPR003006; I9\_MHC.  
DR Pfam: PF00041; I9\_3; 6.  
DR Pfam: PF00047; I9\_8.  
DR SMART: SM00660; FN3\_6.  
DR SMART: SM00409; I9\_9.  
DR SMART: SM00408; I9\_C2\_9.  
DR SMART: SM00410; I9\_Like\_6.  
KW Immunoglobulin domain.  
SQ SEQUENCE 2013 AA; 22268 MW; 1F4AF09A4BBE8A77 CRC64;

Query Match 13.3%; Score 172; DB 11; Length 2013;  
Best Local Similarity 29.8%; Pred. No. 4.2e-06;  
Matches 57; Conservative 33; Mismatches 67; Indels 34; Gaps 10;

QY 15 LPEMTSSVLTNFKTNTYILDTTPSSASLIQAVONHILHEELLMTYKEEYVHAKSKNKI 74  
DB 320 PRKVKSSV-----GSVSLSCSVGN EDDELWYR-NGEI-LNPGKNV 360  
QY 75 -----SSSVVCSSISENDNGISFTGLRGRQDSV---VSVALNTFPPLLSGNDFQVTE 126  
DB 361 KILGLHMANLIMHMKVSKSG-ATCYVRNKLASQVDVYVGLFGIPKILISAFSEKYS 419  
QY 127 EGSNVTI VNVVANPQAMMYKNSSLDLEKSPHQI22--TSES--FQI SITVEKPD 181  
DB 420 EAEVSLVNVNKGITPLEFTWLDLDDPI LKSGGKRISQUMIISKVNVSYLNISSQVRD 478  
QY 182 NGTYSCTAKSS 192  
DB 479 GCYVRCCTANNS 489

RESULT 6  
ID Q07153 PRELIMINARY: PRT: 946 AA.  
AC Q07153;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE RECEPTOR TYROSINE KINASE.  
OS Torpedo californica (Pacific electric ray).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
OC Elasmobranchii; Squali; Hypnosquali; Pristiogadidae; Batoidae;  
OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.

OX NCBI\_TaxID=7787;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC TISSUE-ELECTRIC ORGAN, AND MUSCLE;  
 RX MEDLINE=9321931; PubMed=88549;  
 RA Jennings G.B., Dyer S.M., Burdon S.J.;  
 RT "Muscle-specific tdk related receptor with a kringle domain defines a  
 distinct class of receptor tyrosine kinases";  
 RI Proc. Natl. Acad. Sci. U.S.A. 90:2895-2899(1993).  
 DR EMBL: U11311; AAC9285.1;  
 DR HSSP: P1162; ITCR  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR000024; Fz\_domain.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR InterPro: IPR003600; Ig\_Like.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR000001; Kringle  
 DR InterPro: IPR001245; Tyr\_kinase  
 DR Pfam: PF01392; Fz\_1.  
 DR Pfam: PF00047; Fz\_3.  
 DR Pfam: PF00069; Kringlet; 1.  
 DR Pfam: PF00069; Kringlet; 1.  
 DR PRINTS: PR00018; KRINGLE.  
 DR PRINTS: PR00199; TYRKINASE.  
 DR SMART: SM00408; IG\_C2; 2.  
 DR SMART: SM00410; IG\_Like; 1.  
 DR SMART: SM00130; KR; 1.  
 DR SMART: SM00219; TYRKT; 1.  
 DR PROSITE: PS00348; FZ\_1.  
 DR PROSITE: PS00021; KRINGE\_1; FALSE\_NEG.  
 DR PROSITE: PS00070; KRINGE\_2; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding, immunoglobulin domain, kinase, transferase;  
 KM Tyrosine-protein kinase  
 SQ STQDKMKE 946 AA; 105802 MW: 81A9131DB7419001 CRC64:  
 50  
 Query Match 13.28; Score 171.5; DB 13; Length 946;  
 Best Local Similarity 25.18; Pred. No. 1.8e 06;  
 Matches 63; Conservative 44; Mismatches 95; Indels 55; Gaps 10;  
 OY 7 ::::VIT: FLIR EPTSSVTVVYKTEHY:LTTFGQASLTCVNZHTEF 55  
 DB 10 LLMFLVTTGSAAGTIPKAPQITSPLEFYDAVF-----PEASPCVADSP-VA 59  
 OY 56 ELIAYREGRV--DIKSNKINSSVCVSSISENENG:SFQRLGRDQSVSVVLNV 111  
 DB 60 EITWTRNNITFIRPDTRYSIKEMQILITISVADLVNYYCIIANNQMSAASGALGV 119  
 OY 112 TTPP :LSSNDPQVTHSSNVALVGVKANIQAOMMYKNSSLIDLEKSNIGVQTS 169  
 DB 120 KMKIKLIRPDIVKAL-IVSNVVLKSTMGKPKAISWENQALIKDQPKISV---L 175  
 OY 170 FQSLTKVKEKQVQVTSQIAKSSIAKFE-----SLQHLIVK 205  
 DB 176 GNIIRKVVQLEQKQKCIAPNSLDFEYPSAALEVQVSAKIVAPTSJNVSVSRVILQ 235  
 OY 206 DKTVGVPIEP 216  
 DB 236 GKATGPIEPT 246  
 RESULT 7  
 ID Q9V6C2 PRELIMINARY: PRT: 359 AA.  
 AC Q9V6C2;  
 DE 01-MAY-2000 (TRENKLE: 13, created)  
 DE 01-MAY-2000 (TRENKLE: 13, last sequence update)  
 DE 01-SEP-2001 (TRENKLE: 19, last annotation update)  
 DE LAG PROTEIN (L033460P).  
 GN LAG OR G312369.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota, Metazoa, Arthropoda, Tracheata, Insecta;  
 OC Eukaryota, Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;  
 OC Ephyrotidae, Drosophilidae, Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.F., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Calle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Stiller G.C., Wortman J.R., Yandell M.D., Zhang Y., Chen L.X.,  
 RA Braden R.C., Rogers Y.H.C., Blazer G.G., Chapple M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter K.G., Helt G., Nelson C.R., Mikos G.L.G.,  
 RA Artl J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley S.,  
 RA Beeson K.Y., Benos P.V., Bertan R.P., Bhattacharya D., Bolstad S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burks K.C., Busan D.A., Butler H., Cadien E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
 RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferrar C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iodwam C.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kimminich J.A., Ketchum K.A.,  
 RA Jinnai B.E., Kodira C.D., Frit T., Rivett S., Kulp E., Lai Z.,  
 RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Y., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklow G., Mishina N.V., Mobarry C., Morris J.J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacich J.M.,  
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reiner T., Penington K., Sanders R.D.C., Scheetler F., Shen H.,  
 RA Shue B.C., Siden Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun F.,  
 RA Svirskas R., Teclor C., Turner R., Ventor E., Wang A.H., Wang X.,  
 RA Weng Z., Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh K.F., Zaveri J.S., Zhao M., Zhang S., Zhao Y., Zheng L.,  
 RA Zheng X.B., Zhou F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RI Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN-T, CN BW SP;  
 RA Stapleton M., Brokstein P., Hong L., Agayani A., Carlson J.,  
 RA Chapple M., Chavez C., Daisett V., Farfan D., Fiske R., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mundall C.,  
 RA Nino T., Pacich J., Paragay V., Park S., Phouanavong S., Wan K.,  
 RA Ye C., Lewis S.E., Rubin G.M., Celinker S.;  
 PL Submitted (Aug 2001) to the EMBL/GenBank/TrEMBL databases.  
 DR EMBL: AF003822; AAF38506.1;  
 DR EMBL: AY051829; AAF93253.1;  
 DR HSSP: P56276; ITCR.  
 DR PfamBase: PF00010238; Lac.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR InterPro: IPR003600; Ig\_Like.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam: PF00047; Ig\_3.  
 DR SMART: SM00408; IG\_C2; 2.  
 DR SMART: SM00410; IG\_Like; 1.  
 KW Immunoglobulin domain.  
 SQ STQDKMKE 359 AA; 35935 MW: 124513E24W915044 CRC64:  
 50  
 Query Match 13.18; Score 169.5; DB 5; Length 359;  
 Best Local Similarity 27.18; Pred. No. 8.4e-07;  
 Matches 56; Conservative 42; Mismatches 74; Indels 31; Gaps 10;

```

QY 8 LVYLFLPREMTSVLVNGKTEYLDTTPGQASGLICAVONHTREPELLMYR-EGRY 66
DB 14 LLAIFVQOTLAORTPTISYITOEQIKDI--GGVIEFDCSVQ-YAKENYVLFLEKTDSPV 70
QY 67 DLKSGNKI-----NSSS--VCVSSISENDNGISFTCR--LGRQGSVSVYL 109
DB 71 FLSTGSLVYKRSRSLRDPNPSSTYKLOIKDIOETDAG-TYTCOVVSYVHKYSAEYKL 129
QY 110 MYTPEPLLSGNDPQYV--EEGSSNVLGVNKNAPCAVMMYK-NSSLJLEKSKHQIYQI 106
DB 130 SVRRPPIVSDNSTQGSVVASGESEVMECYASGYPPPTITWRRENNAILPTDGA-----T 183
QY 167 SESFOLSTIKVEKPKDNGTYSCTAKSSL 193
DB 184 YVGNLTIRIKSVKKEKRGTYCVADNGV 210

RESULT 8
Q9NY24 PRELIMINARY: PRT: 431 AA.
AC Q9NY24;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIGLEC SAF2.
GN SAF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kikly K.K., Bochner B.S., Freeman S.D., Tan K.B., Gallagher K.T.,
RA D'Alessio K.J., Holmes S.D., Abrahamson J.A., Erickson-Miller C.L.,
RA Murdoch P.R., Tachimoto H., Schleimer R.P., White J.R.;
RT Identification of SAF-2, a novel single expressed on eosinophils,
RT mast cells and basophils.*
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DB EMBL: AF223403; AAF34702.1;
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003598; Ig_G2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00408; IgC2; 1.
DR SMART: SM00410; Ig_Like; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
KW Immunoglobulin domain.
SQ SEQUENCE 431 AA; 46433 MW; 49B91B8ACE67B118 CkC64;

Query Match 13.1%; Score 169.5; DB 4; Length 431;
Best Local Similarity 23.6%; Pred. No. 1,1e-06;
Matches 63; Conservative 53; Mismatches 92; Indels 59; Gaps 11;

QY 16 REMTSSVLVNGKTEYLDT--TPGQASGLICAVONHTREPELLMYREGEVDLS--GN 72
DB 144 KOLSVVATLTHRPDLILGTLGLESGRNLTSVP-----WACKOGTPPMISWGA 194
QY 73 KINS-----SSVCVSSISENDNGISFTCR--LGRQGSVSVYLWTEPP----- 115
DB 195 SVSSPPPTARRSVLTLTPKPDHGSILTCQVLTGCTVTTSTYRLDVSYPNNLTMTV 254
QY 116 -----LSCGNDPQYVEGSSNVLGVNKNAPCAVMMYKNSSLJLEKSKHQIYQI 166
DB 255 PQGDATASRALNGSSLSLSEGOGLPIYAVANSNPARELSWTPGS--ITTPSP-----S 307
QY 167 SESFOLSTIKVEKPKDNGTYSCTAKSSLTDFHLIVKDTGV--PIEPIIACV--- 221
DB 308 SNGGLELPRVHVRDEGFTCRQANVAGSOSHSLSLQNEGTGSRPVSGVTLAAVGA 367
QY 222 -----VIFLFLCFTLIRKKIKMLCK 244

```

```

DB 368 GATATLFLSCIIIFI-----IVKSCRK 389

RESULT 9
Q921P2 PRELIMINARY: PRT: 605 AA.
AC Q921P2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:12077).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DB EMBL: BC011310; AAH11310.1;
SQ SEQUENCE 605 AA; 67353 MW; DBD98C64D245A867 CkC64;

Query Match 12.8%; Score 166; DB 11; Length 605;
Best Local Similarity 28.0%; Pred. No. 5.2e-05;
Matches 58; Conservative 35; Mismatches 80; Indels 34; Gaps 9;

QY 41 QASLIC-----AVONHTREPELLMYREGEVDLSGKNKINSVCSSISE 86
DB 325 VVPLNEASGUPILPSLIWTSIRNLISSEKTL---DGHMVKSHAV--SLILASLIY 378
QY 87 NDNGISFTC-----ELGRQGSVSVLVNTPPRLSGNDPQYVEGSSNVLGVNKNAPQ 142
DB 379 TQAG-EYICTASNTIGD--SQSMYLEVQYAPKIQGVAAVYTWEGNVCITCEFAVPS 434
QY 143 AUMMWKNSSLDLDELK-SPHYIGTFSFSLIHKVFNW:YVSSIAKSLKFLSLATH 201
DB 435 ATISWRDQGLLPSSNYSNRIKYNTPSASYLEVTTDSNDNGNCTAVNRIGQESLEFI 494
QY 202 LIVKDKTVGVP---IEPIIACVYIF 224
DB 495 LVQAD-IPSSPSIDKVEYSIAVQVF 520

RESULT 10
Q24273 PRELIMINARY: PRT: 1011 AA.
AC Q24273;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NEUROMUSCULIN.
GN NRM OR CG8779.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyridiidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-9400831; PubMed-8398154;
RA Kanla A., Han P.L., Kim Y.T., Bellen H.;
RT "Neuroscultin, a Drosophila gene expressed in peripheral neuronal
RT precursors and muscles, encodes a cell adhesion molecule.*"
RL Neuron 11:673-687(1993).
DB EMBL: L23146; AAA03750.1;
DB FlyBase: FBgn0005629; nrm.
DR InterPro: IPR003599; Ig_.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_9.
DR SMART: SM00409; Ig; 4.

```

DR SMART: SM00410; IC\_1like: 2.  
 KM Cell adhesion.  
 SQ SCQUENCE 1011 AA; 11125 MW; AFWAALAD53AE63 CR664.

Query Match 12.4% Score 161; DB 5; Length 1011.  
 Host local Similarity 24.3% Pred. No. 1.6e-05;  
 Matches 46; Conservative 34; Mismatches 81; Indels 28; Gaps 6;

DB 23 LTVNCKENTYIDTPGQASLIGAVNTPREFELMP-----EGRVDL----- 68  
 ID IDINCYKHH-----TVGSSVVLTCIHGAAPAVNLTWNTTTTSSGNEITFVRSKSL 320  
 QY 69 KSNKINSSSVYSSISFENGTSTFC-----RLSPQSVSVVLTVPPLISND 121  
 DB 421 KSNKTHLSSELENAIRHENAIVFGLAHANIVQJINRNPISLSALTELELVAVKSP 380  
 QY 122 FGVFVPSNKKV-CNFKANICA--GMWYKNSLIDFKSHHQIQDTSFVSTFVE 178  
 DB 481 SATTANTSEVLNCEYFNPASTTQVEWYRNDLVNNDTHYKGSNVALVKSTE 440

QY 179 KPNGTYS 187  
 DB 441 KEDIGNYC 449

RESULT 11  
 QYVNP2 PRELIMINARY PKI: 1079 AA.

AC QYVNP2 PRELIMINARY PKI: 1079 AA.  
 DT 01-MAY-2000 (TREMblrel, 13, last sequence update)  
 DT 01-DEC-2001 (TREMblrel, 19, last annotation update)  
 DE NEM PROTEIN.  
 GN NEM OR C88779.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Phyllophaga; Drosophilidae; Drosophila.  
 NX NCBI\_TaxID:7227;  
 RN [1]  
 KC SEQUENCE FROM N.A.  
 KC STRAIN:BERKELEY;  
 RX MEDLINE:20196006; PubMed:10731132;  
 RA Adams M.D., Gindler S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blake J.C., Champ M., Pfeiffer B.D.,  
 RA Wan K.H., Boyle G., Baxter K.G., Holt G., Nelson C.R., Miklos G.L.G.,  
 RA Arl J.J., Aghayani A., An H.J., Andrews Pankoch C., Baldwin P.,  
 RA Ballow R.M., Basu A., Baxendale J., Bayraktarolu L., Beasley E.M.,  
 RA Boeson K.Y., Bonos P.V., Berman H.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra T.,  
 RA Cherry J.M., Cawley S., Dahlke G., Davenport L.B., Davies P.,  
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Fosden K., Gump L.E., James M., Dugan-Rocha S., Dunkov B.G., Dunn P.,  
 RA Faurin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,  
 RA Host D.D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jallil M., Kalish F., Kafan G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel R.F., Kodira C.D., Kraft C., Kravitz S., Kulp U., Lai Z.,  
 RA Laske P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Morkulov G., Mishing N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Murphy K., Murphy L., Muzny D.M., Nelson J.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rajbani K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue H.C., Siden Klamis I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun F.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng D.,  
 RA Zhou X.H., Zhou F.N., Zhou W., Zhou M., Zhou S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003599; AAF51887.1;  
 DR Flybase: FBgn0005629; nrm.  
 DR InterPro: IPR003599; 19.  
 DR InterPro: IPR003600; 19\_1like.  
 DR InterPro: IPR003006; 19\_MHC.  
 DR Pfam: PF00047; 19; 9.  
 DR SMART: SM00409; IG\_4.  
 DR SMART: SM00410; IC\_1like: 2.  
 SQ QYVNP2 1079 AA; 11125 MW; AFWAALAD53AE63 CR664;

Query Match 12.4% Score 161; DB 5; Length 1079;  
 Host local Similarity 24.3% Pred. No. 1.6e-05;  
 Matches 46; Conservative 34; Mismatches 81; Indels 28; Gaps 6;

DB 23 LTVNCKENTYIDTPGQASLIGAVNTPREFELMP-----EGRVDL----- 68  
 DB 265 IDINCYKHH-----TVGSSVVLTCIHGAAPAVNLTWNTTTTSSGNEITFVRSKSL 320  
 QY 69 KSNKINSSSVYSSISFENGTSTFC-----RLSPQSVSVVLTVPPLISND 121  
 DB 421 KSNKTHLSSELENAIRHENAIVFGLAHANIVQJINRNPISLSALTELELVAVKSP 380  
 QY 122 FGVFVPSNKKV-CNFKANICA--GMWYKNSLIDFKSHHQIQDTSFVSTFVE 178  
 DB 481 SATTANTSEVLNCEYFNPASTTQVEWYRNDLVNNDTHYKGSNVALVKSTE 440

QY 179 KPNGTYS 187  
 DB 441 KEDIGNYC 449

RESULT 12  
 QYVNP2 PRELIMINARY PKI: 597 AA.

AC QYVNP2 PRELIMINARY PKI: 597 AA.  
 DT 01-DEC-2001 (TREMblrel, 19, last sequence update)  
 DT 01-DEC-2001 (TREMblrel, 19, last annotation update)  
 DE STALIC ACID BINDING LECTIN SIGLEC-L1.  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;  
 OC Platyrrhini; Eulipotyrrhini; Primates; Catarrhini; Hominoidea; Pan.  
 NX NCBI\_TaxID:9598;  
 RN [1]  
 KC SEQUENCE FROM N.A.  
 KC MEDLINE:21523976; PubMed:11546777;  
 RA Angata T., Varki N.M., Varki A.;  
 RT "A second uniquely human mutation affecting stialic acid biology.";  
 SL J. Biol. Chem. 276:40282-40287(2001).  
 DR EMBL: AF293372; AAL09302.1;  
 KM lectin.  
 SQ SCQUENCE 597 AA; 65075 MW; 74C13CFAVACDB5BA5 CR664;

Query Match 12.4% Score 159; DB 6; Length 597;  
 Host local Similarity 23.2% Pred. No. 1.3e-05;  
 Matches 63; Conservative 45; Mismatches 85; Indels 78; Gaps 12;

DB 31 NYLID-----TTPGQAS-----11GAVNNTPPFRTI WYPPGRVRLKS 70  
 DB 260 NYLIDTSVVALTHLPFTSPGLTSGHPNLTGSP-----WAGDGPPTIT 310  
 QY 71 GNKINSSSVYSSIS-----ENDNGISFTCR-----GRDSVSVSVLNVTPPP 115

```

Db 311 WNGASVSSLEPTIRSSMLSLPKPOHGTSLTCQVTLPGAGVTTTSAVRINISYPONT 370
OY 116 -----LLEGNDFTEVEGSSNVKLCVKNKANPQUMMMYNNSSLLDEKRRHQ 162
Db 371 TMTVPOGDOCTASTLTARNSALSYLEGSLHVCADVSPARPLSMTGOS--LTLSPS--- 425
OY 163 IQOTSBSFOLSTIKVEKPDNCTYSCIASSLKTESLDFHLIVKDTYGV--PIEPI---- 216
Db 426 --QSSMIGVLELPRVHVDEGEFTCRQNPGLSQHISLSLSQNETYTKMKMPSISVTLGA 483
OY 217 ---IACVIFELTFCGLIARRKKIMKCMK 244
Db 484 VGGAGATALVPLSFCTITFV-----VRSCKR 509

RESULT 13
O55124 PRELIMINARY: PRT: 1463 AA.
AC O55124:
DT 01-JUN-1998 (TREMBLrel 06, Created)
DT 01-JUN-1998 (TREMBLrel 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel 19, Last annotation update)
DE M-PROTEIN.
GN MYOM2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=98234547; PubMed=9570952;
RA Steiner F., Weber K., Fuerst D.O.;
RT *Structure and expression of the gene encoding murine M-protein, a
RT sarcomere-specific member of the immunoglobulin superfamily.*;
RL Genomics 49:83-95(1998).
DR EMBL: AJ001038; CAA0492.1;
DR HSP: P56276; IRLK.
DR MGD: MGI:1328358; Myom2.
DR InterPro: IPR003962; FNIII_repeat
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003598; Iq_C2.
DR InterPro: IPR003600; Iq_1like.
DR InterPro: IPR003006; Iq_MHC.
DR Pfam: PF00041; fn3; 4.
DR Pfam: PF00047; Ig; 4.
DR PRINTS: PR00014; FNTPETII.
DR SMART: SM00060; FN3; 4.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00410; Iq_1like; 2.
DR Immunoglobulin domain; Repeat.
SQ SEQUENCE 1463 AA; 164518 MW; 8C6551ED90F0B288 CRC64;

Query Match 12.2%; Score 158; DB 11; Length 1463;
Best Local Similarity 25.1%; Pred. No. 4,7e-05;
Matches 60; Conservative 36; Mismatches 97; Indels 46; Gaps 10;

OY 21 SVLVNKGKT-ENVIL--DTFHSUAS-----LTCAYVNHREBELLYREGR 65
Db 1211 SVLEVGKGYEDMILLAMSVGCASASPLKVLCTPEGTLQCFMKYFTEEMKVSMTHEKAK 1270
OY 66 VDLKSGKNINSSS---VCVSSISENDNG-ISFTGRLGRD----- 100
Db 1271 ISSSEHMRIGSEEMAMWLQICEPTEKDKGYTFEFDCKSHQSHSLDLSGQAFDEAAEF 1330
OY 101 QSVSVSVVNTVTFPPLLS-NDQFTEBGSNVKLCVKNKANPQUMMMYNNSSLLDEKS 159
Db 1331 QDLKAAPAEKRGKGVIGLIPVVTIMGKTLNLCTVFGNPDDEYVFFKNDK--DIELS 1388
OY 160 RHQIQOTSSFOLST--KVEKPDNCTYSCIASSLKTESLDFHLIVKDTYGV-IEP 215
Db 1389 EHFIVAMBSQKYSVLTIGCVTAEDSGKSLNVKNKYGGEKIDVTVSYTKHGEKIPVLS 1447

```

```

RESULT 14
O94155 PRELIMINARY: PRT: 1415 AA.
AC O94155:
DT 01-FEB-1997 (TREMBLrel 02, Created)
DT 01-FEB-1997 (TREMBLrel 02, Last sequence update)
DT 01-DEC-2001 (TREMBLrel 19, Last annotation update)
DE UNC-40 (T19B4.6 PROTEIN).
GN UNC-40 OR T19B4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL;
RA Chan S.S.Y., Zheng H., Su M.W., Wilk R., Killen M.T., Hedgecock E.M.,
RA Cuiotti J.G.;
PL Submitted (SEP-1996) to the FMR1/GenBank/DBJ databases.
RN 12
RP SEQUENCE OF 1248-1415 FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnecough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Garner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Yaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome 111 of C.
RT elegans.*;
RL Nature 368:32-38(1994).
RN 13
RP SEQUENCE OF 1248-1415 FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gattung S.;
KL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN 14
RP SEQUENCE OF 1248-1415 FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U70618; AAB17088.1;
DR EMBL: U80438; AAB37634.1;
DR HSP: P40189; IBOU.
DR InterPro: IPR003962; FNIII_repeat.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003598; Iq_C2.
DR InterPro: IPR003600; Iq_1like.
DR InterPro: IPR003006; Iq_MHC.
DR Pfam: PF00041; fn3; 4.
DR Pfam: PF00047; Ig; 4.
DR PRINTS: PR00014; FNTPETII.
DR SMART: SM00060; FN3; 6.
DR SMART: SM00408; IGC2; 3.
DR SMART: SM00410; Iq_1like; 1.
DR Immunoglobulin domain; Repeat.
SQ SEQUENCE 1415 AA; 154436 MW; B2D90F76F9FEB6A9 CRC64;

Query Match 12.2%; Score 157.5; DB 5; Length 1415;
Best Local Similarity 27.8%; Pred. No. 5e-05;
Matches 58; Conservative 35; Mismatches 83; Indels 33; Gaps 10;

OY 16 REMTSSVLVNGKT-ENVYIDTP-----GSGASLCAVNGHTREBELLYREGR- 65
Db 247 KSSQTAKLTVTEIVSNELVITTPRIQVYVGGUDEFLECLVASLR-PQVKKLKDSKQI 285

```

66 - VILKSKKINSSSVSSISNMGTSFTPL--DROSVSVVLNVTPEPLLSGNPQ 124  
 286 IVDGVIHKRVSSILVSBASHLEPGL-VTCASNNIUSLURASVEXKAPRILIKR 144  
 124 TVE-EESNRYKVC-NVKANPOAKMRYAN-----SSILDEKSHOJUSHSKUS 175  
 445 KAVETAIVELRGTAAAPARVMWKNCEALIGSEFEVLEPRN-----LRIL 393  
 176 KVEKPKNGTYSCIASLAKTESIDPLILV 204  
 394 GYVRAIAVYCTAFENDVSEQASAILV 422  
 RESULT 15  
 09VPO8 PRELIMINARY: PRT: 467 AA.  
 01-MAY-2000 (TREMble, 13, Created)  
 01-MAY-2000 (TREMble, 13, Last sequence update)  
 01-DEC-2001 (TREMble, 19, Last annotation update)  
 CG7166, PROTEIN.  
 Drosophila melanogaster (fruit fly).  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID:7227;  
 111  
 SEQUENCE FROM N.A.  
 STRAIN-BERKELEY;  
 MEDLINE:20196006; PubMed:10741132;  
 Adams M.D., Celisner S.R., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amaralides F.G., Scherer S.E., Li P.W., Hoskins R.A., Gill R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pretter B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Adney A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,  
 Boonch K.Y., Bonos P.V., Borker B.P., Bhandari D., Bolshakov S.,  
 Borovaya D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
 Burks K.C., Busam D.A., Butler H., Caden E., Caceres A., Chandra T.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 Glodok A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,  
 Hartis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,  
 Jalili M., Kalish J., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimel H.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Metkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy T., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 Palazov M., Pittman G.S., Pan S., Pohlard J., Part V., Reese M.G.,  
 Rehorst K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svrtkars R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Woynick G.M., Weissbach J.,  
 Williams S.M., Woodage T., Wortley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaccari J.S., Zhu M., Zhang G., Zhao Q., Zheng L.,  
 Zhou X.H., Zhou F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 \*The genome sequence of Drosophila melanogaster.\*;  
 Science 287:2185-2195(2000).  
 R. EMBL: AF004595; AAF51754.1;  
 D. Flybase: FBgn0037107; CG7166.  
 D. InterPro: IPR001064; Crystalin.

DR InterPro: IPR003598; 19\_c2.  
 DP InterPro: IPR003600; 19\_1like.  
 DP InterPro: IPR003605; 19\_MHC.  
 UK Pfam: PF00047; 19f\_3.  
 DR SMART: SM00408; 19c2; 2.  
 DR SMART: SM00410; 19\_1like; 1.  
 DR ProSITE: PS00226; CRYSTALLIN, PFTAGAMMA; UNKNOWN\_1.  
 KW Immunoglobulin domain.  
 SO SEQUENCE 467 AA; 53023 MW; 5D0DCFC9GDD3BH47E CRC64;

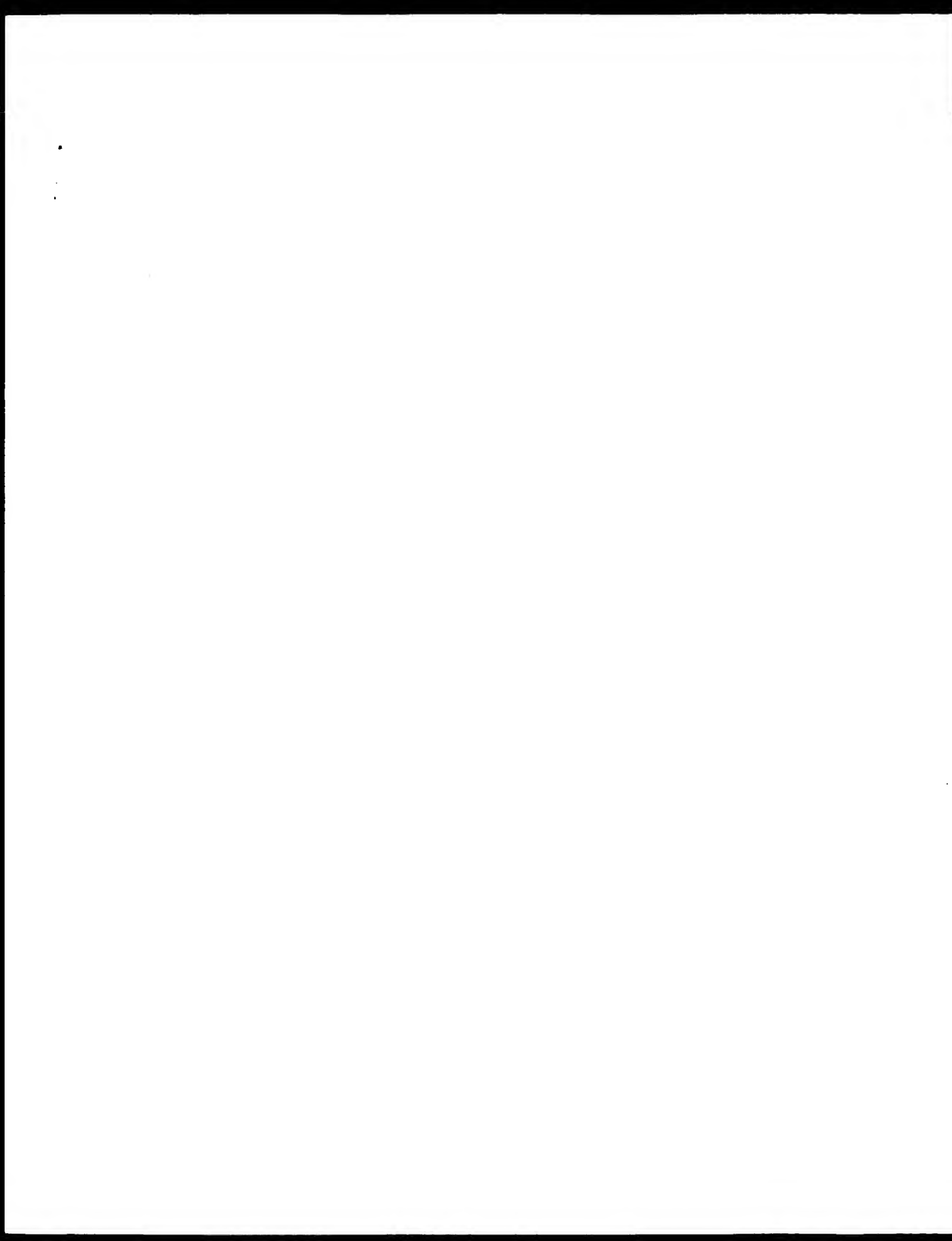
Query Match 12.1%; Score 157; DB 5; Length 467;  
 Best Local Similarity 26.5%; Pred. No. 1.4e-05;  
 Matches 48; Conservative 37; Mismatches 72; Indels 24; Gaps 8;

24 TVNCKTEHYLDFTTQSGASLIGAVGNFKEKELMVEKRYVLSK--NKINSSVCV 81  
 136 TURALPHNGVATARKGSIVLECKASNP-VPIIPWRK-----DPSGPIHLSISSTIL 190  
 82 SSTSEPMWTSFTPL--DROSVSVVLNVTPEPLLSGNPQVE-----EESNRYK 132  
 191 EAVDSHHGS-TYQGSADNGVADRYVSMIDGILTSIPPE-----TVKSWVHASEYDVE 243  
 133 LVGVNVAIVCAQMMWKNSSILDEKSHOJUSHSKUS 192  
 244 LVGVNVAIVCAQMMWKNSSILDEKSHOJUSHSKUS 192  
 193 L 193  
 303 L 303

Search completed: September 17, 2002, 13:23:22  
 Job time: 804 sec











DB 186 APPSATISWFRDQGLSSSNYSNIKINYTPSASYLEVTPUSENDFGNYNCCTAVNFIQUES 245  
 QY 206 LDFHLIVKDKTVGPV----IEPIAACVIF 232  
 DB 246 LEFLIVQAD-TPSSPSIDQVEFYSIAVOVF 275

RESULT 4  
 ID P91255 PRELIMINARY, PRT, 2762 AA.  
 AC P91255:  
 DT 01-MAY-1997 (TREMblrel, 03, Created)  
 DT 01-DEC-2001 (TREMblrel, 19, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel, 19, Last annotation update)  
 DE HYPOTHEICAL 306.5 KDA PROTEIN.  
 GN F12F3.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL NZ;  
 RX MEDLINE=98069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.\*;  
 RL Science 282:2012-2018(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL NZ;  
 RA Fulton B., Wohlmann P.;  
 RT "The sequence of C. elegans cosmid F12F3.\*;  
 RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL NZ;  
 RA Waterston R.;  
 RT "Direct Submission.\*;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U80022; AAC25886.2; -;  
 KM Hypothetical protein.  
 SQ SEQUENCE 2762 AA; 306525 MW; C964CCEH6D322084 CRC64;

Query Match 13.0%; Score 173.5; DB 5; Length 2762;  
 Best Local Similarity 24.1%; Pred. No. 5e-06;  
 Matches 60; Conservative 36; Mismatches 100; Indels 53; Gaps 9,

QY 12 GFFLLVILFPREMTSVLTNGKTEYILDTTP-----GS 48  
 DB 1562 GFNFCLINEGQASASQCVTFIFKPPAS--LQSTPDHSLERNLVPTLOKALNNESSAQGO 1619  
 QY 49 GASLCAVQNHITREELIWPDE-----GVDLKSNGKINSSVCSSLSNDNGISFT 102  
 DB 1620 QIMTITRISRS-ESTVAMFPRDEPIESAGPTFISNRKFSNHLVLC-HAVQSDPTG-KTP 1676  
 QY 103 C----RIGRDS--VSYSVVLNVTFPPLISGNDPOTVEEGSNKLYCNKANPOAMM 155  
 DB 1677 CVVTKKYGAASECNVAVEDYTKFIAPSFATLSDSITALLGHNTLECKVEGSPAPEVSW 1736  
 QY 156 YKNSSLILKESRHOITQOTSFPOLSTTKVKKPKPMGTYSCTAKS-----SLKT 203  
 DB 1737 TKDGRISTTRRIROTQDENCNCLSLSKAESIDMVGYYVTSVAGVSTSSNMVIAKT 1795  
 QY 204 ESLDFHLIV 212  
 DB 1797 TGTDSHLVI 1805

RESULT 5  
 Q9V6C2  
 Q9V6C2

ID Q9V6C2 PRELIMINARY, PRT, 359 AA.  
 AC Q9V6C2:  
 DT 01-MAY-2000 (TREMblrel, 13, Created)  
 DT 01-MAY-2000 (TREMblrel, 13, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel, 19, Last annotation update)  
 DE LAC PROTEIN (LD33460P).  
 GN LAC OR CG12369.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Urosophillidae; Urosophila.  
 OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu T., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport I.B., Davies P.,  
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong P., Gorrell T.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.P., Hong J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegun G.,  
 RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson R.P., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weislock G.M., Weissenbach J.,  
 RA Williams S.M., Wudage I., Worley K.C., Wu D., Yang S., Ye Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang S., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.\*;  
 RL Science 287:2185-2195(2000).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Y, CN BW SP;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Paclob J., Paragas V., Park S., Phouanrenvong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RT Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE003822; AAF98506.1; -;  
 DR EMBL: AY051829; AAK93253.1; -;  
 DR HSSP: p56276; ITLK.  
 DR FlyBase: FBgn010238; LAC.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR InterPro: IPR003600; Ig\_Like.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam: PF00047; Ig\_3.  
 DR SMART: SM00408; IgC2; 2.  
 DR SMART: SM00410; Ig\_Like; 1.

KW Immunoglobulin domain.  
 SEQ ID: 459 AA: 39949 MW: 125513E2B9C156F4 CRC64:

Query Match 12.9%, Score 172, DH 13, Length 199,  
 Best Local Similarity 27.3%, Pred. No. 4,6e-07,  
 Matches 60, Conservative 43, Mismatches 85, Indels 32, Gaps 11,

UY 4 MASSIVEGQKRLVYVLPDGMSSVAVNKAFTNTLITDQASLLAVNINR 62  
 :  
 DB 2 WPISTNVAWVSLTALATPVQDILAPDIPDIPSYLPEYKDKL--DQVTSVQVYAKR 67  
 :  
 UY 63 EELLWVPECHVDLKSNGK-----NSSV--VAVSSISENINQVLPV-- 104  
 :  
 DB 58 YVNLTKTSQVFLSTSTIVIKISPESTPYDQSSGVYKIQKQDGLAC--FYDQV 116  
 :  
 UY 106 LQHDQSVSVVIVNTPPHLISNIPQTV--PSSNVAITVNVKANDQANMAYK--NSSI 161  
 :  
 DB 117 LSTVHKVSAEVMISVRRPVLSDNSTQSVASRSESEVMQCYASQYPTPTWRENNAL 176  
 :  
 UY 162 LDEKSHQLOOTSPQISTKVKRPNGTYSCTAKSSL 201  
 :  
 DB 177 LPTPSA-----TVGNILKLSVNHLPQIYVVALNIV 210

## RESULT 6

UYERCB PRELIMINARY: PRT: 2013 AA.

ID UYERCB  
 AC UYERCB: 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE DOWN SYNDROME CELL ADHESION MOLECULE.  
 GN DSCAM.  
 OS Mus musculus (Mouse).  
 OC Fungiata; Metazoa; Chordata, Craniata, Vertebrata, Euteleostomi,  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RA Barlow G.M., Lyons G.E., Korenberg J.R.;  
 RT "Down syndrome cell adhesion molecule, a highly conserved mouse  
 homology, descam, is expressed in the neonatal and adult mouse brain.";  
 RT Submitted (oct-2000) to the EMBL/GenBank/DDBJ databases  
 RI 121  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1CR OUT-BRED;  
 RA Agarwala K.L., Banesh S., Amato K., Suzuki T., Yonakawa K.;  
 RT "DSCAM, a highly conserved gene in mammals, expressed in  
 differentiating mouse brain.";  
 RI Biochem. Biophys. Res. Commun. 281:697-705(2001).  
 DR EMBL: AF315558; AAC28796.1;  
 DR EMBL: AY005483; AAF99440.1;  
 DR M30: M31196281; Dscam.  
 DR InterPro: IPR001961; FN\_111.  
 DR InterPro: IPR001594; Iq.  
 DR InterPro: IPR003598; Iq\_C2.  
 DR InterPro: IPR003600; Iq\_Like.  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR Pfam: PF00041; Iq3; 6.  
 DR Pfam: PF00047; Iq; 8.  
 DR SMART: SM00060; FN3; 6.  
 DR SMART: SM00409; Iq; 9.  
 DR SMART: SM00408; IqC2; 9.  
 DR SMART: SM00410; IqLike; 6.  
 KW Immunoglobulin domain.  
 SEQ ID: 2013 AA: 22268 MW: 144AF07947BFA77 CRC64:

Query Match 12.9%, Score 172, DH 13, Length 2013,  
 Best Local Similarity 29.8%, Pred. No. 4,6e-06;

Matches 57, Conservative 33, Mismatches 67, Indels 34, Gaps 10;

UY 23 PREMLSVLVNPKLNTLITDQASLLAVNINR 82  
 :  
 DB 320 PREYSSV-----DSVSSISVYGN--ELVLSWY--NPT--INDKRV 360  
 :  
 UY 83 -----NSSSVVSSISNDNSTLPTCEGRVSV--SVVNVTPPLSGNDQVTE 144  
 :  
 DB 451 PELLGIMNMIIMHGVNSLQGTAPVYVKIKLSAQDYVGVILHGLPILISAEVYS 419  
 :  
 UY 145 PLSNVAIVNVKANDQANMAYKNSV--PSSNVAITVNVKANDQANMAYK--NSSI 189  
 :  
 DB 420 PAVPVSVNVNPTVPTMTLITDQV--KSNHPLSMILISGVNVSLNLSNVQVRO 478  
 :  
 UY 160 NQIYSCTAKSSL 200  
 :  
 DB 479 GAVVPTAVNS 489

## RESULT 7

ID 007153 PRELIMINARY: PRT: 946 AA.

AC 007153  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE RECEPTOR TYROSINE KINASE.  
 OS Torpedo californica (Pacific electric ray).  
 OC Eukaryota; Metazoa; Chordata, Craniata, Vertebrata, Chondrichthyes;  
 CC Torpediniformes; Torpedinidae; Torpedinidae; Torpedo.  
 OX NCBI\_TaxID=7787;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=ELECTRIC ORGAN, AND MUSCLE;  
 RA Jemima C.G.B., Dyer S.M., Burden S.J.;  
 RT "Muscle-specific trk related receptor with a kringle domain defines a  
 RT distinct class of receptor tyrosine kinases.";  
 RI Proc. Natl. Acad. Sci. U.S.A. 90:2895-2899(1993).  
 DR EMBL: L11311; AAA49285.1;  
 DR HSSP: P11362; IFGK.  
 DR InterPro: IPR003719; FcK\_kinase.  
 DR InterPro: IPR000024; Fc\_domain.  
 DR InterPro: IPR003598; Iq\_C2.  
 DR InterPro: IPR003600; Iq\_Like.  
 DR InterPro: IPR003006; Iq\_MHC.  
 DR InterPro: IPR000001; Kringle.  
 DR InterPro: IPR01345; Tyr\_kinase.  
 DR Pfam: PF01392; F2; 1.  
 DR Pfam: PF00047; Iq; 3.  
 DR Pfam: PF00051; Kringle; 1.  
 DR Pfam: PF00059; PKinase; 1.  
 DR PRINTS: PR00018; KRINGLE.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR SMART: SM00408; IqC2; 2.  
 DR SMART: SM00410; IqLike; 1.  
 DR SMART: SM00109; KR; 1.  
 DR SMART: SM00219; TYRK; 1.  
 DR PROSITE: PS0038; F2; 1.  
 DR PROSITE: PS00021; KRINGLE\_1; FALSE\_NEG.  
 DR PROSITE: PS00070; KRINGLE\_2; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Immunoglobulin domain; Kinase; Transferase;  
 KW Tyrosine-protein kinase.  
 SEQ ID: 946 AA: 105892 MW: 81A9131DB7479D01 CRC64:

Query Match 12.8%, Score 171.5, DH 13, Length 946;  
 Best Local Similarity 25.1%, Pred. No. 2e-05;  
 Matches 63, Conservative 48, Mismatches 95, Indels 55, Gaps 10;

[illegible]

RESULT	8			
Q9NYZ4				
ID	Q9NYZ4	PRELIMINARY:	PRT:	431 AA.
AC	Q9NYZ4:			
DT	01-OCT-2000 (TREMblrefl, 15, Created)			
DT	01-OCT-2000 (TREMblrefl, 15, Last sequence update)			
DT	01-DEC-2001 (TREMblrefl, 19, Last annotation update)			
DE	SIGLEC SAF2.			
GN	SAF2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.			
OX	MBRL_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Kikly K.K., Bochner B.S., Freeman S.D., Tan K.B., Gallagher K.T.,			
RA	D'Alessio K.J., Holmes S.D., Abrahamson J.A., Erickson-Miller C.L,			
RA	Murdoch P.R., Tachimoto H., Schleimer R.P., White J.R.;			
RT	"Identification of SAF-2, a novel siglec expressed on eosinophils			
RT	mast cells and basophils.";			
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF223403; AAF34702.1; ..			
DR	InterPro: IPR001899; Gram_Pos_anchor.			
DR	InterPro: IPR003598; Ig_C2.			
DR	InterPro: IPR003600; Ig_1like			
DR	InterPro: IPR003006; Ig_MHC.			
DR	Pfam: PF00047; Ig_2			
DR	SMART: SMD0408; IGC2; 1.			
DR	SMART: SMD0410; IGC1; 1.			
DR	PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.			
DR	Immunoglobulin domain.			
CO	SEQUENCE 431 AA: 46433 MW: 49D91B84CE67B118 CRC64;			

	Query Match	12.7%	Score 169.5;	DB 4,	Length 431;
	Best Local Similarity	23.6%;	Pred. No. 1.2e-06;		
	Matches	63;	Conservative	53;	Mismatches 92; Indels 59; Gaps 11
QY	24	REMSVLYTNGKTENYLDT--TPGQSALICAVONHRTREELMYREEGRVDLKS--CN 80			
	:	:::	: ::	::	::
Db	144	KQLSFVTATLTNRPDLLILGLTGLESGRNLTCSP-----WACKQGTPPMTSMIGA 194			
QY	81	KINS-----SSVCASSISENDNGISFICRL---GROOSVSYSVLNWTFRP----- 123			
	:	:::	: ::	::	::
Db	195	SVSSPGFTARRASSTLTLPKPODHGSLSCOVTLGTGTSTTSIVRKLDSVPKPNLTMIV 254			
QY	124	-----ILLSDNFQVEEGSNVLCNVANCAQAMMMYNKNSLLDEKSHOLUOT 174			
	:	:::	: ::	::	::
Db	255	FQGGATASTALGNSSLSYLEGSQLRYCAVANSPPALISMTRGS--LTLDPSR-----S 307			
QY	175	SESQQLSIITVEKRPDNCYTSCIARKSLKTESLDPHLIVKDITGVG--PIEPIIAACY-- 229			

Lb 308 SNGGELLEKRVHAKRGGELICAGNANUSQHIHLSLSLSLNEBIOGIRKRVSVYLAANSGA 367  
 QY 230 ----VIFILQCFGLIAPRKIMKLMK 252  
 Qb 368 GATATAFLSPCTTFT-----VPSPPK 389

RESULT	9	
Q921P2		
Q921P2	PRELIMINARY:	PT: 605 AA.
AC	Q921P2	
DT	01-DEC-2001 (TREMBLE, 19, Created)	
DT	01-DEC-2001 (TREMBLE, 19, Last sequence update)	
DT	01-DEC-2001 (TREMBLE, 19, Last annotation update)	
DE	UNKNOWN (PROTEIN FOR MGC:12077).	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
OX	NCBI_TaxID=10090;	
RN	[1]_TaxID=10090;	
RA	SEQUENCE FROM N.A.	
RA	Strausberg R.	
KL	Submitted (JUL-2001) to the EMBL/GenBank/DBS databases.	
DR	EMBL: BC011310; AF011310.1; ..	
SO	SEQUENCE: 605 AA 67343 MW. CPH9664D245A67 CPH64:	

Query Match	12.48;	Score 166;	DB 11;	Length 605;
Best Local Similarity	28.08;	Pred. No. 3.5e-06;		
Matches 58;	Conservative 35;	Mismatches 80;	Indels 34;	Gaps 9;

```

QY 49 QASLTIC-----AONMTRFEELWYREGGRDLSKIMSSVCASSTSE 94
      | : : : : : : : : : : : : : : : : : : : : : : : :
Db 325 QVLTICEASGDPLPSTWTRTSTRNLSSEKTL-----DGHMYVRSARV--SULTIKSTQY 378
      | : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 NDNGISFTC----RGRDQSVSVYVLNVTTPPLLSGNDFOYVEGSSNYKLYCVKANPQ 150
      | : : : : : : : : : : : : : : : : : : : : : : : :
Db 379 TDAG-ETICASMTIGQD---SQSMYLEVGYAPRLQGPVAVYTWBEGNOINICEVFAAPS 434
      | : : : : : : : : : : : : : : : : : : : : : : : :
QY 151 AQUMHMKKNSLLDLK-SKHULUOTISHSFUSLILKVAHPLNGTYSLAKSLIKITSLULH 205
      | : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 ATISWFEDGOLLFPSSNYNSIKIYNTFPSASYLEVPESENDPCFNCTAVNRIGDSELEFI 494
      | : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 LIYEDKTVGVP-----LEPIAACVVI 232
      | : : : : : : : : : : : : : : : : : : : : : : : :
Db 495 LVQAD-TFSSPSIDVEVEPSSIAVQVH 520
      | : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
024273
ID 024273 PRELIMINARY: PR: 1011 AA.
AC 024273:
DT 01-NOV-1996 (TREMblrel: 01, Created)
DT 01-NOV-1996 (TREMblrel: 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel: 19, Last annotation update)
DE NEUROMUSCULIN.
NR OR CG8779.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha.
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9400831; PubMed=8198154.
RA Kaita A., Han P.L., Kim Y.T., Bellen H.;
RT "Neuromusculin, a Drosophila gene expressed in peripheral neuronal
RT precursors and muscles, encodes a cell adhesion molecule.";
PI Neuron 11:673-687(1993)
DR EMBL: L23146; AAA03750.1; -.
DR FlyBase: FBgn0005629; ctm.
DR Interpro: IPR003599; Iq.

```

```

DB InterPro: IPR004600; 19_Like.
DB InterPro: IPR004006; 19_MHC.
DB Pfam: PF00047; 19_9.
DB SMART: SM00409; 19_4.
DB SMART: SM00410; 19_Like; 2.
DB Cell address.
DB SOURCE: 1011 AA; 11475 MW; APPRISAP01P3APV63 CPG64;

Query Match
Best Local Similarity 24.38; Score 161, DB 5, Length 1079;
Matches 46; Conservative 34; Mismatches 81; Indels 28; Gaps 6;

31 LTNGKTEHYITDTPGQASLIGAVQNTREELLWYR-----EGGVLD----- 76
DB 265 IDINGVKHH-----TVGSKVVLTEHICAPAVNLTWNTTIISSGNEILEVSKSLK 320
QY 77 KSNCKINSSVCSSISENNGISFTG-----RIGRQSVSVSVLVNTPPILSND 129
DB 121 KSDGTFHUSLITFNAIRFNDVRFCAANIVLQINRNPISALTEVLAPVAVKVSF 380
QY 140 FGVVFPSNKKVY-TVKANPQQA-GMMYKNSLITLTKSRHQIQQLSEPSLSLVE 186
DB 381 SAITANTSELVLNCFYFANPASTGVWYRNLIIVNNDTHYKGSSENVAVIKSTE 440
QY 187 KPNNGTYSQ 195
DB 441 KEDIGNYSQ 449

RESULT 11
Q95LH0: PRELIMINARY; PRT; 1079 AA
AC Q95LH0:
DT 01-MAY-2001 (TrEMBLrel. 19, Created)
DT 01-MAY-2000 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DB NIM PROTEIN.
CN NIM OR C08779.
OS Dirosophila melanogaster (Fruit Fly).
OC Pharyngata; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Dirosophiliidae; Dirosophila.
OX NCBI_TaxID=7227;

RA [1]
RA SEQUENCE FROM N.A.
RA STRAIN-BERKELEY.
RA MEDLINE: 20196006; PubMed 10741132,
RA Adams M.D., Gumbel S.E., Holt P.A., Evans C.A., Gonyea J.D.,
RA Adamides P.G., Scherer S.R., Li P.W., Hoskins K.A., Galle R.R.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Smith G.G., Wortman J.R., Yandell M.D., Zhang Y., Chen L.X.,
RA Brandon R.C., Roberts Y.H., Blazer V., Chapple M., Pfeiffer H.D.,
RA Won K.H., Doyle G., Baxter R.G., Holt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Abmayy A., An H.J., Andrews Flannick C., Baldwin D.,
RA Holow R.M., Hsu A., Hakendale J., Bayraktarolu L., Hensley E.M.,
RA Iversen K.Y., Bonos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Harris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Fahlke G., Davenport L.B., Davies P.,
RA de Paulis R., Delcher A., Deng Z., Myers A.P., New I., Gietz S.M.,
RA Iodson K., Jupp L.E., Jones M., Jughan-Kocha S., Junkov R.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong P., Gorrell J.B., Gu Z., Guan P., Harris M.,
RA Harris M.L., Harvey D., Helman T.J., Hernandez J.F., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibenwale C.,
RA Jolani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
RA Kimmel H.B., Kodira C.D., Kall C., Kivitz S., Kulp D., Lai Y.,
RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Morkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy H., Murphy L., Muzny D.M., Nelson D.L.,

```

```

RA Nelson D.L., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun F.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang C.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhou X.H., Zhou F.N., Zhou W., Zhou X., Zhu G., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Dirosophila melanogaster."
RL Science 287:2185-2195(2000).
DB EMBL: AE003599; AAF51887.1;
DB FLYBASE: FBgn0005629; nrm.
DB InterPro: IPR003499; 19_
DB InterPro: IPR003499; 19_Like.
DB InterPro: IPR003006; 19_MHC.
DB SMART: SM00409; 19_4.
DB SMART: SM00410; 19_Like; 2.
DB SOURCE: 1079 AA; 121454 MW; C75658224C027B CPG64;

Query Match
Best Local Similarity 24.38; Score 161, DB 5, Length 1079;
Matches 46; Conservative 34; Mismatches 81; Indels 28; Gaps 6;

31 LTNGKTEHYITDTPGQASLIGAVQNTREELLWYR-----EGGVLD----- 76
DB 265 IDINGVKHH-----TVGSKVVLTEHICAPAVNLTWNTTIISSGNEILEVSKSLK 320
QY 77 KSNCKINSSVCSSISENNGISFTG-----RIGRQSVSVSVLVNTPPILSND 129
DB 121 KSDGTFHUSLITFNAIRFNDVRFCAANIVLQINRNPISALTEVLAPVAVKVSF 380
QY 140 FGVVFPSNKKVY-TVKANPQQA-GMMYKNSLITLTKSRHQIQQLSEPSLSLVE 186
DB 381 SAITANTSELVLNCFYFANPASTGVWYRNLIIVNNDTHYKGSSENVAVIKSTE 440
QY 187 KPNNGTYSQ 195
DB 441 KEDIGNYSQ 449

RESULT 12
Q95LH0: PRELIMINARY; PRT; 597 AA.
AC Q95LH0:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DB STALIC ACID-BINDING LECILIN s16GEC-11.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
OX NCBI_TaxID=9596;
RA [1]
RA SEQUENCE FROM N.A.
RA MEDLINE: 21523976; PubMed 11546777;
RA Argata T., Varli N.M., Varli A.;
RA "A second uniquely human mutation affecting stalic acid biology."
RL J. Biol. Chem. 276:40282-40287(2001).
DB EMBL: AE293372; MAF09302.1;
KM Lectin.
DB SOURCE: 597 AA; 65075 MW; 74C13CFACD85BA5 CPG64;

Query Match
Best Local Similarity 23.28; Score 159, DB 6, Length 597;
Matches 63; Conservative 45; Mismatches 85; Indels 78; Gaps 12;

39 NYILD-----TTPGQAS-----LICAVQNTREELLWYRFGKVIDKS 78

```



```

EL 260 NYIVDK:SVHVTALHTLFTSFCTLESCHPRLNLCSSV: MACFCGHTTIL 310
OY 79 --GKINSSSVYCVASIS-----ENDNGISFPCRL---CRQSVSVSVLVNTFPP--- 123
Db 311 WMGASVSSLEPTISRSSMLSLIKRPODHGTSITCQVTLTPGAGVTTTAAALINISTEPOL 370
OY 124 -----LLSSNDPJVLEBJSNKLVCNFKANPVAMWYKNSLLULEKSRU 170
Db 371 TMTVFOGDGTAFTLRNGSALSVLEGOSLHVCADSNPPARLSWTGWS--LTLSPS-- 425
OY 171 IQQTSFQSLSTKVEKPMQKTSCLAKSLKTESLDFHLIVDKTVGV--PIEPI----- 224
Db 426 --QSSNLGVLELPRVHAKDEGEFCRAONPLGSOHISLSLSDQNEYTGKMRPLSGVTLGA 483
OY 225 --IACGVIFLTLCFL:ARRKKIMKCMK 252
Db 484 VGGAGATATLVFLSFCITFV-----VVRSCRK 509

```

```

RESULT 13
ID 055124 PRELIMINARY: PRT: 1463 AA.
AC 055124:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE M-PROTEIN.
GN MYO2.
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
CC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98214547; PubMed=9570952;
RA Steiner F., Weber K., Fuerst D.O.;
RT "Structure and expression of the gene encoding murine M-protein, a
RT sarcomere-specific member of the immunoglobulin superfamily.";
RL Genomics 49:83-95(1998).
DR EMBL: AJ001038; CAA04492.1; -.
DR HSSP: P56276; 1TLK.
DR MCD: MGI:1328358; Myom2.
DR InterPro: IPR003962; FN.II repeat.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003598; Ig_L2.
DR InterPro: IPR003600; Ig_Like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Ig; 4.
DR PRINTS: PR00014; FNTYPELIT.
DR SMART: SM00060; FN3; 4.
DR SMART: SM00408; IGC2; 1.
DR SMART: SM00410; IG_Like; 2.
KW Immunoglobulin domain; Repeat.
SQ SEQUENCE 1463 AA; 164518 MW; 8C6551ED90F0B288 CRC64;

```

Query Match 11.8%; Score 158; DB 11; Length 1463;  
 Best Local Similarity 25.18; Pred. No. 5; 2e-05;  
 Matches 60; Conservative 36; Mismatches 97; Indels 46; Gaps 10;

```

OY 29 SVLTATNGKT-RNYIL-DTFPGSQAS-----LICAVQNHTRREELLMTYEEDR 73
Db 1211 SVLEGGGVYEDMLTAMSRVCGASAPLKVICTPEGRITLCGFKKFTFEKKNVWYKKEK 1270
OY 74 VDLKSGKNTSSS-----VCVSSISENDG--ISFTCLGRD----- 108
Db 1271 ISSSEHMRIGGSEEMAMLIQICEPTEKDKGYTFEIFDGKDSHORSLDLSQAFDEAYAEF 1330
OY 109 QSVSVSVLVNTFPPLLSG-NDPQVEGGSNVKVCNFKANPVAMWYKNSLLULEKSRU 167
Db 1331 QQLKAAAFKAEKRNKGVIGGLPDVYITMEGKTLNLTCTVFCNPDPEVYVWFKNDX--DIEIS 1388

```

```

OY 168 NKLQVLELSLQSL:KWKILDNQYVSSIANSSLRKTESLPLHLIVKQATVGV 223
Db 1989 PHIVKMEGSKYVSLTIGYTATPDSKMYKSYINVKNKYGTREKIDVTAVVYHCKRIDISF 1447

```

```

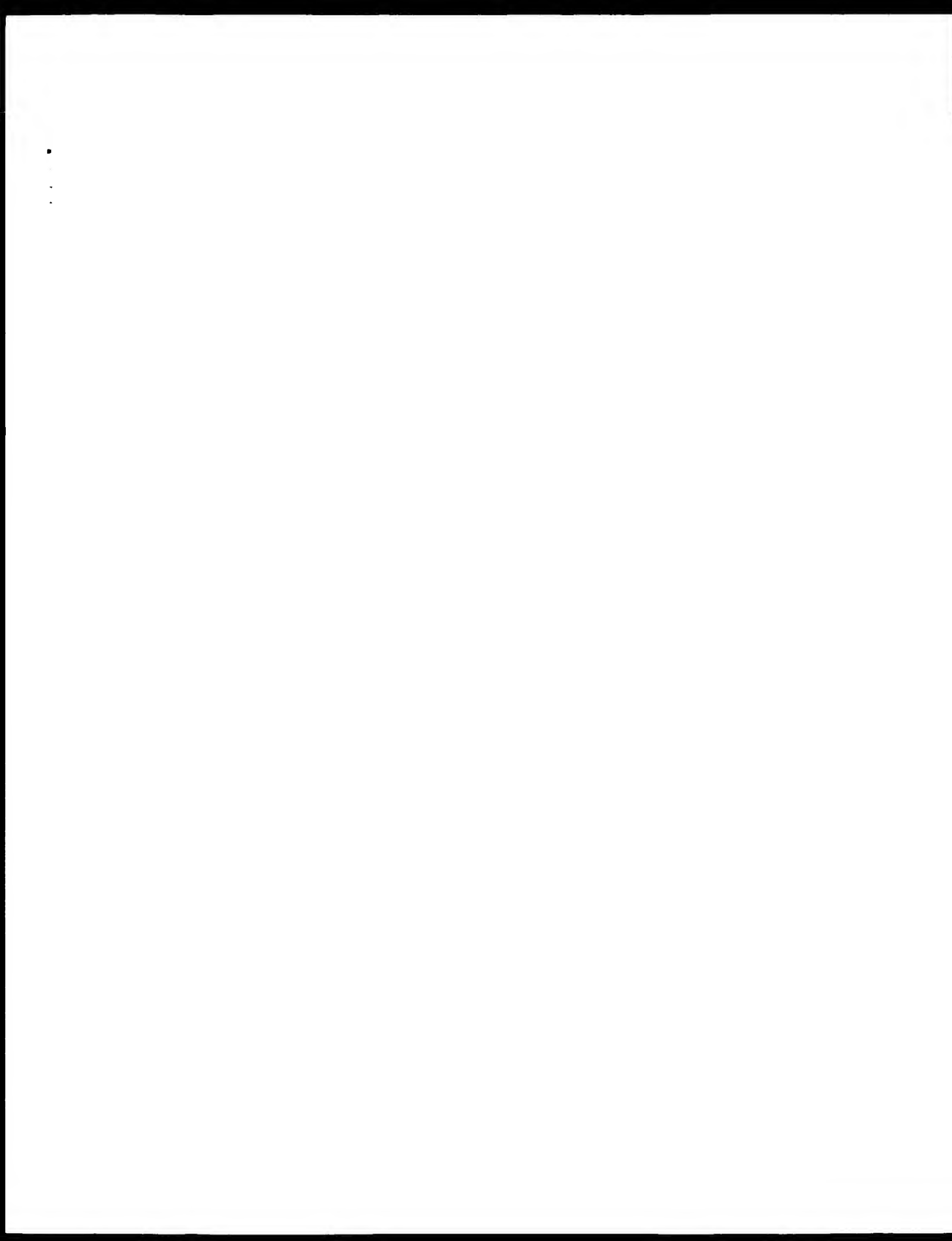
RESULT 14
ID 094155 PRELIMINARY: PRT: 1415 AA.
AC 094155:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE UNC-40 (P1984.6 PROTEIN).
GN UNC-40 OR P1984.6.
OS Caenorhabditis elegans.
OC Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditioidea:
CC Rhabditidae: Pelodierinae: Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL;
RA Chan S.S.Y., Zheng H., Su M.W., Wilk R., Killen M.T., Hedgecock E.M.,
RA Cuiotti J.G.;
RL Submitted (SEP-1996) to the FMR1/GenBank/DBD databases.
RN [2]
RP SEQUENCE OF 1248-1415 FROM N.A.
RC STRAIN-BRISTOL N2;
RA MEDLINE=94150718; PubMed=7905798;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Boufield J., Burton J., Cunnell M., Cussey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Moutrey A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Riffen L., Poopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson Spiro J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE OF 1248-1415 FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gattung S.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBD databases.
RN [4]
RP SEQUENCE OF 1248-1415 FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBD databases.
DR EMBL: U70618; AAB17088.1; -.
DR EMBL: U80438; AAB37634.1; -.
DR HSSP: P40189; 1HOU.
DR InterPro: IPR003962; FN.II repeat.
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR003598; FN.II.
DR InterPro: IPR003600; Ig_L2.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00041; fn3; 6.
DR Pfam: PF00047; Ig; 4.
DR PRINTS: PR00014; FNTYPELIT.
DR SMART: SM00060; FN3; 6.
DR SMART: SM00408; IGC2; 3.
DR SMART: SM00410; IG_Like; 1.
KW Immunoglobulin domain; Repeat.
SQ SPQFWNE 1415 AA; 154476 MW; R2Q9QF74FQ9FR5A9 CRC64;

```

Query Match 11.8%; Score 157.5; DB 5; Length 1415;  
 Best Local Similarity 27.88; Pred. No. 5; 5e-05;  
 Matches 58; Conservative 35; Mismatches 83; Indels 33; Gaps 10;







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 10:01:52 ; Search time 375.67 Seconds

Title: US-09-863-823-5

Sequence: 1 atgcatggaagcagtgt.....acagtgaacagctctatga 789

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters. 3472872

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match	08
Maximum Match	100%

Database : N\_Geneset\_032802 \*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT *
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT *
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT *
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT *
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT *
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1965.DAT *
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT *
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT *
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT *
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT *
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT *
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT *
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT *
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT *
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT *
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT *
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT *
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT *
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT *
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT *
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT *
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT *
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT *
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARY

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	558.4	70.8	1018	22	AA528788	Human Immunoglobulin
c 2	441	55.9	1958	22	AAH35000	Human IgG1 gamma 1
c 3	284	36.0	19315	22	AA528883	Human Immunoglobulin
c 4	277.4	35.2	338	22	AA537585	Novel human diacytin
5	250.8	31.8	361	22	AA548121	Novel human diacytin
6	249	31.6	358	22	AA538280	Novel human diacytin
7	247.2	31.3	359	22	AA565873	Novel human polynucleotide
8	203.4	25.8	359	22	AA565873	Novel human polynucleotide
9	41.6	5.3	2949	21	AA000524	Degenerate DNA entom

[illegible]



CC mutation in a DNA sequence or determining the presence or amount of  
 CC expression of the protein. Alternatively the identification of a binding  
 CC partner to a sequence allows determination of changes in protein  
 CC activity. The sequences can be used as research tools for receptors or  
 CC other signal transduction pathway proteins that interact with the  
 CC polypeptides of the invention and can be used to treat, prevent or  
 CC diagnose various types of disorders such as neurological disorders,  
 CC cardiovascular disorders, gastrointestinal disorders, reproductive  
 CC disorders, immune system disorders, renal disorders, muscular disorders,  
 CC pulmonary disorders, proliferative disorders and cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from Wipo  
 CC at ftp.wipo.int/pub/published\_pcr\_sequences.

XX Sequence 1018 BP; 264 A; 225 C; 198 G; 311 T; 0 other:

Query Match 70.8%; Score 558.4; DB 22; Length 1018;  
 Best Local Similarity 99.1%; Pred. No. 1.2e-164;  
 Matches 559; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 79 acaagttctgtttaaactgtgaatgaactgaactatatactgatactacact 138  
 DB 963 AGAGTTCTGTTTAACTGTGAATGTAATAACTGAGAACTATATCTGATACACACT 904  
 OY 139 ggtcccaagcctctctgtagtctgttcaaaaccacacagagagagaaactgtc 198  
 DB 903 GGTCCCAAGCATCTCTGATATGCTGTCTTCAAAACCAACCAAGAGAGAAATGCTC 844  
 OY 199 tgtaccagagagagagagagagattgaatctgaactgaactgaactgaact 258  
 DB 843 TGTACCGAGAGAGAGAGAGAGATTGAATCTGGAACCAAAATCAATTCAGACT 784  
 OY 259 gttctgtctcttccatcagtgagaaatgacaaggaactgaactgaactgaact 318  
 DB 783 GTCTGTCTCTTCCATCAGTGAATGACCAACGCAATCAGCTTACCTGAGCTGAGG 724  
 OY 319 aggaatcagtcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgtctcgt 378  
 DB 723 AGGAGATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 664  
 OY 379 ggaagagacttccaaacagtgagagagagagagagagagagagagagagag 438  
 DB 663 GGAAGAGACTTCCAAACATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 604  
 OY 439 gcaaacacacag 498  
 DB 603 GCCAACCCAGGCTCAAAATGATGCTGACAAACACAGTACCTGATTTAGAGAA 544  
 OY 499 agcagtcac 558  
 DB 543 AGCCGTCACCAATTCACACAGACAGACAGTGTCTTTTCACTGTCAATACCAAGTCA 484  
 OY 559 aagcctgacacaggaactacagttgattgcaagtcacatctctgaagagagag 618  
 DB 483 AACCGTACACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424  
 OY 619 gacttccactgattgattaaat 642  
 DB 423 GACTTCACTGATTTGTTAACT 400

## RESULT 2

AAH35000 standard; cDNA: 1958 BP.

XX AAH35000.

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding cDNA SEQ ID NO:2082.

XX Human; colon cancer; colon cancer antigen, diagnosis, detection.

KW colorectal carcinoma; ss.  
 XX Homo sapiens.  
 OS  
 XX WO200122920-A2.  
 PN  
 XX 05-APR-2001.  
 PD  
 XX 28-SEP-2000; 2000WD-US26524.  
 PE  
 XX 29-SEP-1999; 99US-0157137.  
 PR 03-NOV-1999; 99US-0164280.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
 PI  
 XX Wt1, 2001-23345724.  
 UK  
 XX P-PSDB; AAG75595.  
 DR  
 XX Nucleic acids encoding 4277 human colon cancer associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 XX Claim 1; Page 3568-3569; 9803pp; English.

CC AAH32943 to AAH37195 and AAG77788 represent human colon  
 CC cancer associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patient's own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated P,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAG77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N: H Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SKU ID No: 1027 to 1052, 7921 and 7922.

XX Sequence 1958 BP; 509 A; 460 C; 462 G; 495 T; 12 other:

Query Match 55.9%; Score 441; DB 22; Length 1958;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-127;  
 Matches 441; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 349 ctgaatgllactlltct 408  
 DB 644 ctgaatgllactlltct 703  
 OY 409 agtaatgtaaatggttggtaaatgtaaatgtaaatgtaaatgtaaatgtaaatg 468  
 DB 704 agtaatgtaaatggttggtaaatgtaaatgtaaatgtaaatgtaaatgtaaatg 763  
 OY 469 aaaaacagtagtct 528  
 DB 764 aaaaacagtagtct 823  
 OY 523 tctttcagctgtcattcaccacaaagtcgagaaagcttgacacaggaactaaagttgatt 588  
 DB 824 tctttcagctgtcattcaccacaaagtcgagaaagcttgacacaggaactaaagttgatt 883  
 OY 589 gcaaatcatctctgaaacaggaagagagagagagagagagagagagagagagag 648  
 DB 884 gcaaatcatctctgaaacaggaagagagagagagagagagagagagagagagag 943  
 OY 649 gtaggtgatacagatagagcccatattgtgcatgtgtgtgtgtgtgtgtgtgtgtgt 708







XX  
-

```

Sequence 328 BP; 102 A; 62 C; 81 G; 83 T; 0 other;
Query Match:      31.4%; Score 249; EB 22; Length 328;
Best Local Similarity 97.6%; Pred. No. 7 8e-68;
Matches 249; Conservative 3; Mismatches 3; Indels 0; Gaps 0

```

QY 1 atggcatggaagagcagtgatcatalatgaggaagatttctcttaagtaattta 60  
 |||  
 Db 73 atggcatggaagagcagtgatcatalatgaggaagatttctcttaagtaattta 132  
 QY 61 ttctgcacgtgagatgacaaagtctgttttaactgtgaatgataaactgagaactat 120  
 |||  
 Db 133 ttctgcacgtgagatgacaaagtctgttttaactgtgaatgataaactgagaactat 192  
 QY 121 atccctgatactacacccgtgctcccaagcatctctgataatgtgtcttcaaacacac 180  
 |||  
 Db 193 atccctgatactacacccgtgctcccaagcatctctgataatgtgtcttcaaacacac 252  
 QY 181 aagaaggaagaaactgtctgtgtacccgagagagagagatggaatttgaactgtsaac 240  
 |||  
 Db 253 aagaaggaagaaactgtctgtgtacccgagagagagagatggaatttgaactgtsaac 312  
 QY 241 aamaatcaattccagc 255  
 |||  
 Db 313 aaaaatcaattccagc 327

## RESULT 7

AAF65373

AAF65373 standard; cDNA; 359 BP.

XX AAF65373;

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide; SEQ ID NO: 1129.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;

XX breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX MO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

XX (CHIR ) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

XX Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

XX Crtkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

XX Kita D, Garcia V, Jones LM, Strache-Clrain B;

XX WPI; 2001-091805/10.

XX Library of polynucleotides for diagnosing a cancerous state of a

XX mammalian cell and detecting cancer, particularly of the colon or

XX prostate, comprises 3351 human polynucleotide sequences -

XX Claim 9: Page 706; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human  
 CC polynucleotides. The library is used to detect differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell and can  
 CC detect colon, prostate, breast and lung cancer. The library can be used  
 CC to produce probes for detection of mRNA and to produce additional copies  
 CC of the polynucleotides. The probes can be used for chromosome mapping of  
 CC the polynucleotide and for detection of transcription levels. Ribozymes  
 CC or antisense oligonucleotides can be generated. The polynucleotides and  
 CC their gene products are used as genetic or biochemical markers (e.g. in  
 CC blood or tissues) that will detect the earliest changes along the  
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and  
 CC preventive interventions. The polynucleotides, polypeptides and

CC antibodies against them can be used in pharmaceutical compositions to  
 CC treat the cancers and proliferative disorders such as neoplasia,  
 CC dysplasia and hyperplasia.

XX Sequence 359 BP; 113 A; 64 C; 92 G; 89 T; 1 other;

Query Match 31.3%; Score 247.2; DB 22; Length 359;

Best Local Similarity 98.4%; Pred. No. 3e-67;

Matches 246; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggcatggaagagcagtgatcatalatgaggaagatttctcttaagtaattta 60  
 |||  
 Db 109 atggcatggaagagcagtgatcatalatgaggaagatttctcttaagtaattta 168  
 QY 61 ttctgcacgtgagatgacaaagtctgttttaactgtgaatgataaactgagaactat 120  
 |||  
 Db 169 ttctgcacgtgagatgacaaagtctgttttaactgtgaatgataaactgagaactat 228  
 QY 121 atccctgatactacacccgtgctcccaagcatctctgataatgtgtcttcaaacacac 180  
 |||  
 Db 229 atccctgatactacacccgtgctcccaagcatctctgataatgtgtcttcaaacacac 288  
 QY 181 aagaaggaagaaactgtctgtgtacccgagagagagagatggaatttgaactgtsaac 240  
 |||  
 Db 289 aagaaggaagaaactgtctgtgtacccgagagagagagatggaatttgaactgtsaac 348  
 QY 241 aamaatcaattccagc 250  
 |||  
 Db 349 aaaaatcaattccagc 358

## RESULT 8

AAF65369

AAF65369 standard; cDNA; 359 BP.

XX AAF65369;

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide; SEQ ID NO: 1125.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;

XX breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX MO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US18374.

XX 02-JUL-1999; 99US-0142310.

XX 02-JUL-1999; 99US-0142311.

XX (CHIR ) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

XX Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

XX Crtkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

XX Kita D, Garcia V, Jones LM, Strache-Clrain B;

XX WPI; 2001-091805/10.

CC Library of polynucleotides for diagnosing a cancerous state of a  
 CC mammalian cell and detecting cancer, particularly of the colon or  
 CC prostate, comprises 3351 human polynucleotide sequences -  
 CC Claim 9: Page 706; 1046pp; English.  
 CC The present sequence is one of 3351 sequences in a library of human

50 Sequence 559 BP; 106 A; 69 C; 95 G; 88 T; 1 other;

[illegible]

by the fact that the  $\mathcal{L}_1$  norm is not differentiable at the origin. In the case of the  $\mathcal{L}_2$  norm, the solution is given by the least squares solution, which is unique. In the case of the  $\mathcal{L}_1$  norm, the solution is given by the least absolute values solution, which is not unique. In the case of the  $\mathcal{L}_\infty$  norm, the solution is given by the least maximum solution, which is not unique. In the case of the  $\mathcal{L}_p$  norm, the solution is given by the least  $p$ -norm solution, which is not unique. In the case of the  $\mathcal{L}_1$  norm, the solution is given by the least absolute values solution, which is not unique. In the case of the  $\mathcal{L}_2$  norm, the solution is given by the least squares solution, which is unique. In the case of the  $\mathcal{L}_\infty$  norm, the solution is given by the least maximum solution, which is not unique. In the case of the  $\mathcal{L}_p$  norm, the solution is given by the least  $p$ -norm solution, which is not unique.

00 241 00 242

AAID(00)21

XX  
DT 24-JUL-2000 (11rst enlrv)

KW Chronic inflammatory demyelinating polyneuropathy; contact dermatitis;

PN 971270076, 1V 971270076, W02000216, A1,

[illegible]

xx  
xx  
xx

Conklin DC; Ellsworth TJ;

P1 CONKLIN DE, ELLSWORTH JL;  
XX

22 The present sequence encodes follistatin related protein 216a2. This is

CC chronic inflammatory demyelinating polyneuropathy, various forms of  
CT contact dermatitis, contact vulvovaginitis, mycositis, scrofula and

Sequence 2949 BF; 614 A; 274 C; 477 G; 396 T; 1188 other;

GV 239 acbbaabbaagctcttacccttgcagagcttggggaaggaatcagctccgtgtccgtttccagatggtc 349

Ob 1472 Lyallghyll, Cretaceous Limestone, near junction of road 1471

DL 1472 y09tuoeyawssytlmjltpyayp6n4azgnaLhcnarCcnarytnqntqy 1531

[illegible]

00 587 +1000000 594

XX  
XX  
AAV70330.

aa Human receptor tyrosine kinase ROR-2 encoding DNA.

KW binding protein; BDNF; NT-3; diagnosis; ss.

key	location/Qualifiers
CD5	200-3031
FT	
FT	

XX  
PN  
1155843749-A

PN  
XX  
055843/49-A.



XX	25- JAN -2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03 AUG-2000; 2000US-0511451.
PR	15-SEP-2000; 2000US-0661870.
XX	
XX	(HSP)- HVSFQ INF
PA	
PI	Tang YT, Liu G, Zhou F, Qian XB, Wang Z, Chen R, Asundi Y,
P1	Cao Y, Duanac RA, Zhang J, Werham TJ
XX	
OR	WP1: 2001-476164/51.
OR	p-PStm; AAM3951.
XX	
PS	Claim 1: Page 509-510; 127pp; English.
XX	
CC	The present invention provides the protein and coding sequences of novel
CC	proteins from a variety of organisms, including human, dog, cat, horse,
CC	cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC	urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC	from the organism of interest. They can be used in diagnostics,
CC	forensics, gene mapping, identification of mutations, to assess
CC	biodiversity and for nutritional purposes. The present sequence is a cDNA
CC	of the invention.
SQ	Sequence 4710-60; 1174 A; 1258 C; 1167 G; 1106 T; 9 other:
Query Match	5.0%; Score 39.42 Db 42; Length 4710;
Best Local Similarity	51.1%; Ered: Not 9.42;
Matches	92; Conservative 0; Mismatches 88; Indels 0; Gaps 0
DY	415 gtaaaatggtttgcgaatcgtgaagcccaaccctcccgagctcaaatgatgttgataaaaaac 474
DB	
DY	836 gtcaacctggtagtcatgagaaggcttcgccgaagcccccacacttgaagtcttgatacaaggat 895
DB	
DY	475 actagttctctctcatttatgagaagaagccttcaacaacttcaaacagacaaagtgaatctlll 534
DB	
DY	535 caattatcaatcaccaaatccagagagccttgaacaggaacctaagttgattatggaaa 594
DB	
DY	956 catctgacctcaaaaagatgatatgaagaaagaaatgaaatgaaatgaaatgaaatgaaatgaaatg 1015
RESULT 14	
AAD14203	
ID	AAD14203 standard; DNA; 8513 bp.
AC	AAD14203;
XX	
DE	MESVp/MSCT/IRISRCAM(roci) vector.
KW	transcriptional regulatory element; translational regulatory element;
KW	gene expression; diagnostic application; therapeutic application;
KW	MESVp; retrovirus vector; enhanced green fluorescent protein; EGFP; IRBS;
KW	nuclear ribosome and its subunit; NEM; nuclear cell adhesion molecule;
KW	ori; origin of replication; SV40; human; mouse; ds.
XX	
CH	Chimeric - Murine embryonic stem cell virus.
OS	Chimeric - fous sarcoma virus.
OS	Chimeric - Mastadenovirus.
OS	Chimeric - Mus sp.
OS	Chimeric - rhesus macaque polyoma virus.
OS	Chimeric - Homo sapiens.
OS	Chimeric - Synthetic.
XX	
MO	2000155371-A1.

XX	02-AUG-2001.
XX	
XX	26-JAN-2001; 2001WO-US92733.
XX	
PR	28-JAN-2000; 2000MS-0179416.
PS	02-MAY-2000; 2000MS-0166406.
PR	30-MAY-2000; 2000MS-0207804.
PR	07-SEP-2000; 2000MS-0235852.
PR	07-SEP-2000; 2000MS-0230656.
PR	12-JAN-2001; 2001US-0261312.
XX	
FA	(Scri ) Scripps Res Inst.
F1	Maurice VP, Chapman GM, Chappell GM, Jones FS, Owens G, Meech R;
DR	WPI; 2001 4B3242/52.
XX	
PT	Identifying oligonucleotides with transcriptional/translational regulatory activity in eukaryotic cells by integrating an oligonucleotide into cell genome and detecting a change in expression of expressible polynucleotides -
PS	Claim 95; Page 131-135; 172pp; English.
XX	
CC	The present invention relates to a method for the identification of an oligonucleotide with transcriptional/translational regulatory activity in a eukaryotic cell. The method involves integrating an oligonucleotide into the genome of a eukaryotic cell so that it is linked to an expressible polynucleotide or contacting the eukaryotic cell with library of vectors, obtained by cloning a library of oligonucleotide into multiple copies of expression vectors comprising polynucleotide, and detecting changes in expression of the polynucleotide. The method is useful for identifying an oligonucleotide having transcriptional or translational activity in a eukaryotic cell. It is also useful for identifying synthetic transcriptional or translational regulatory elements. The transcriptional or translational regulatory elements are useful in a variety of gene expression configurations for regulating control of expression, and in expression vectors for controlling gene expressions in diagnostic and therapeutic applications. The present sequence is MEVSR/EGFP/LHNSCMAR(tori) vector based on murine embryonic stem cell virus (MESV) retrovirus. The vector comprises nucleotide sequences encoding enhanced green fluorescent protein (EGFP) and human neural cell adhesion molecule (R CAM) linked by an internal ribosome entry sequence (IRES). The upstream long terminal repeat (LTR) region of the retrovirus was modified to contain Rous sarcoma virus (RSV) enhancer elements. The downstream ITR u3 region was modified to contain a cassette containing a polylinker for the insertion of random oligonucleotides, adenovirus major late promoter, the initiator sequence (Inr) from mouse terminal deoxynucleotidyl transferase gene and a complete R region. The vector also contains a simian virus 40 (SV40) origin of replication (ori) and is useful for identifying synthetic transcriptional regulatory elements.
XX	
XX	Sequence 8513 BP; 2067 A; 253 C; 224 G; 1869 T; 9 other;
XX	
Query Match	5.0%; Score 39.2; DB 23; Length 8513;
Best Local Similarity	51.1%; Pred. No. 0.29;
Matches	32; Conservative 0; Mismatch 84; Indels 0; Gaps 0
Cy	415 -gttaagtcttgcttgcgaatgagacggcaaccctcgtctccaaatcggtatcagaac 474 
Dd	3607 gttacctgt 3666 
Gy	475 agtagcttcctcgaatttagagaaaagccgcacccaactccacaacataagtgactctt 534 
Dd	3667 ggaggagacdgatlaagcgcagagagagagagagagagagagagagagagagagag 3726 
Gy	535 caagtgttaactaaccaaatctgagaaagcttcgaacaggaacctacagtcttatctgaag 594 
Dd	3727 cagcttgagccctcdaadaaagtgctatgaacagacagagagagagagagagagagag 3786 

RESULT 14  
 AAT90472  
 ID AAT90472 standard: cDNA: 2869 BP.  
 XX  
 AC AAT90472;  
 XX  
 DT 27-JAN-1998 (first entry)  
 XX  
 DE Rat muscle-specific kinase (MUSK) cDNA.  
 XX  
 KW Receptor tyrosine kinase; muscle specific kinase; MUSK; Dmk; rat;  
 KM ligand; agrin; diagnosis; therapy; ss.  
 XX  
 OS Rattus sp.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 135..2741  
 FT /tag= a  
 FT sig\_peptide 135..191  
 FT /tag= b  
 FT mat\_peptide 192..2738  
 FT /tag= c  
 XX  
 PN M09721811-A2.  
 XX  
 PD 19-JUN-1997.  
 XX  
 PF 13-DEC-1996; 96WO-0520696.  
 XX  
 PR 10-MAY-1996; 96US-0644271.  
 PR 15-DEC-1995; 95US-0008657.  
 XX  
 PA (REF-) REGNERON PHARM INC  
 PI Bowen DC, Glass DJ, Valenzuela DM, Yancopoulos GD;  
 DR WPI: 1997-332783/30.  
 DR P-PSDB: AAM26610.  
 XX  
 PT Nucleotide sequences encoding human agrin and muscle specific kinase  
 PT and related receptor - used in diagnosis and treatment of disorder  
 PT with muscle atrophy  
 XX  
 PS Example 1; Fig 1: 120pp; English.  
 CC This cDNA sequence, deposited as ATCC 75498, includes a coding  
 CC sequence for a novel rat receptor tyrosine kinase designated muscle  
 CC specific kinase or MUSK (see AAM26610) that is expressed in normal  
 CC and denervated muscle. MUSK is alternatively referred to Dmk for  
 CC denervated muscle kinase. The cDNA clone was isolated from  
 CC denervated muscle cDNA using primers (see AAT90476 and AAT90480)  
 CC based on tyrosine kinase homology domains. Human MUSK (see  
 CC AAT90473) has also been identified. Use of MUSK to generate anti-  
 CC MUSK antibodies and in the diagnosis of neurological or other  
 CC disorders is disclosed. Assay systems that may be used to detect  
 CC and/or measure ligands that bind the musk gene product are provided.  
 CC A claimed method of promoting the growth, differentiation or  
 CC survival of MUSK receptor-expressing cells involves administration  
 CC to the cell of agrin (see AAM26609). Such cells include muscle,  
 CC heart, spleen, ovary and retina cells, or cells genetically  
 CC engineered to express the MUSK receptor.  
 CC  
 SO Sequence 2869 BP; 727 A; 754 C; 726 G; 658 T; 4 other;

Query Match 4.9%; Score 38.6; DB 18; Length 2869;  
 Best local Similarity 55.3%; Pred. No. 0.25;  
 Matches 99; Conservative 0; Mismatches 74; Indels 6; Gaps 1;

OY 415 gtaagtgttgatgcaatggaagccacccagctcaatgatgtgtgtacaaaac 474  
 DB 267 ggcacttcatgtgcgtggaatcctacccctcagcctgaattctcttgaccagaat 326

RESULT 15  
 AAT87074  
 ID AAT87074 standard: cDNA: 2869 BP.  
 XX  
 AC AAT87074;  
 XX  
 DT 06-JAN-1998 (first entry)  
 XX  
 DE Rat Dmk receptor cDNA.  
 XX  
 KW Dmk receptor; tyrosine kinase receptor; signal transduction; assay;  
 KM therapy; diagnosis; ss.  
 XX  
 OS Rattus sp.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 135..2741  
 FT /tag= a  
 FT sig\_peptide 135..191  
 FT /tag= b  
 FT mat\_peptide 192..2738  
 FT /tag= c  
 XX  
 PN US5656473-A.  
 XX  
 PD 12-APR-1997.  
 XX  
 PF 21-JUL-1993; 93US-0095658.  
 XX  
 PR 19-JAN-1995; 95US-0374834.  
 PR 21-JUL-1993; 93US-0095658.  
 XX  
 PA (PRF-) PRGENFRON PHARM INC.  
 PI Rojas EA, Valenzuela DM;  
 DR WPI: 1997-414593/38.  
 DR P-PSDB: AAM26507.  
 XX  
 PT New isolated human denervated muscle kinase receptor - used to  
 PT develop products for the diagnosis and treatment of neurological,  
 PT muscle or neuromuscular disorders  
 XX  
 PS Example 1; Column 23-28; 31pp; English.  
 CC This cDNA clone, deposited as ATCC 75498, codes for rat Dmk receptor  
 CC (AAM26507), a novel tyrosine kinase receptor that is expressed at high  
 CC levels in denervated muscle. It was isolated from a rat partial  
 CC skeletal muscle cDNA library following isolation of a partial  
 CC sequence using primers (see AAT87075 and AAT87081) based on receptor  
 CC tyrosine kinase (RTK) homology domains. PCR primers (AAT87075-81)  
 CC based on homology domains between novel Dmk and RTKs can be used to  
 CC identify novel RTKs. A human clone (see AAT87073) coding for a  
 CC claimed human Dmk receptor (AAM26506) has been obtained.  
 CC  
 SO Sequence 2869 BP; 727 A; 754 C; 726 G; 658 T; 4 other;

Query Match 4.9%; Score 38.6; DB 18; Length 2869;  
 Best local Similarity 55.3%; Pred. No. 0.25;  
 Matches 99; Conservative 0; Mismatches 74; Indels 6; Gaps 1;

OY 415 gtaagtgttgatgcaatggaagccacccagctcaatgatgtgtgtacaaaac 474













PR 17-NOV-2000; 2000US-0249214  
PR 17-NOV-2000; 2000US-0249214  
PR 17-NOV-2000; 2000US-0249215  
PR 17-NOV-2000; 2000US-0249216  
PR 17-NOV-2000; 2000US-0249217  
PR 17-NOV-2000; 2000US-0249218  
PR 17-NOV-2000; 2000US-0249219  
PR 17-NOV-2000; 2000US-0249224  
PR 17-NOV-2000; 2000US-0249245  
PR 17-NOV-2000; 2000US-0249264  
PR 17-NOV-2000; 2000US-0249265  
PR 17-NOV-2000; 2000US-0249267  
PR 17-NOV-2000; 2000US-0249299  
PR 17-NOV-2000; 2000US-0249300  
PR 17-NOV-2000; 2000US-0250160  
PR 01-DEC-2000; 2000US-0250391  
PR 05-DEC-2000; 2000US-0251030  
PR 05-DEC-2000; 2000US-0251988  
PR 05-DEC-2000; 2000US-0256719  
PR 06-DEC-2000; 2000US-0251479  
PR 08-DEC-2000; 2000US-0251856  
PR 08-DEC-2000; 2000US-0251858  
PR 08-DEC-2000; 2000US-0251869  
PR 08-DEC-2000; 2000US-0251989  
PR 11-DEC-2000; 2000US-0254097  
PR 05-JAN-2001; 2001US-0259678  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Harash SC, Kuppen SM;  
PI WPI: 2001-457725/49.  
XX  
XX Isolated novel immunoglobulin polypeptide for monitoring the presence  
PT and progression of diseases and for diagnosis -  
XX  
XX Claim 1: SEQ ID No 245; 551bp; English.  
XX  
XX Sequences AAS28878-AAS28926 represent genomic DNA molecules which encode  
CC the immunoglobulin polypeptides of the invention. The polynucleotides and  
CC polypeptides can be used to diagnose a pathological condition or a  
CC susceptibility to a pathological condition in a subject by determining  
CC the presence or absence of a mutation in a DNA sequence or determining  
CC the presence or amount of expression of the protein. Alternatively the  
CC identification of a binding partner to a sequence allows determination of  
CC changes in protein activity. The sequences can be used as research tools  
CC for receptors or other signal transduction pathway proteins that interact  
CC with the polypeptides of the invention and can be used to treat, prevent  
CC or diagnose various types of disorders such as neurological disorders,  
CC cardiovascular disorders, gastrointestinal disorders, reproductive  
CC disorders, immune system disorders, renal disorders, muscular disorders,  
CC pulmonary disorders, proliferative disorders and cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPU  
CC at ftp.wip.int/pub/published\_pcr\_sequences.  
XX  
XX Sequence 19315 BP; 5674 A; 4041 C; 4034 G; 5565 T; 1 other;  
SQ  
Query Match 37.1%; Score 284; DB 22; Length 19315;  
Best Local Similarity 91.2%; Pred. No. 2e-78;  
Matches 299; Conservative 2; Mismatches 27; Indels 0; Gaps 0;  
QY 12 aagattctctctgaattattatctgcacgtagagatgaagaattctgtttaac 71  
DB 14845 AGGTGTTACTCTGTATATCTAAATAGTTTCTCCCTATACCCAGAGTTCTGTTTAAAC 14786  
QY 72 tttgaatgtaaacctgagaactatatactgatactacactggtctccagaactctc 131  
DB 14785 TGTGAATGTAACCTGAGAATATATCTGATCTACTACCTGCTCCCAAGCATCTCT 14726  
QY 132 gataatgctgtttcaaacacacacagagagagaaactgtctgttaccgagagagag 191  
|||||

DB 14725 GATATCTCTCTCTTAAATATATATGAAGAGAAATGCTATGCTATAAGAGAGAG 14666  
QY 192 gagagtgagatcttggaatctgaaacaaatcaatccatccatctctctgttctccat 251  
|||||  
DB 14665 GATATCTCTCTCTTAAATATATATGAAGAGAAATGCTATGCTATGCTATAAGAGAGAG 14606  
QY 252 cag 311  
|||||  
DB 14605 CAGTGAATAATCAGAACCGAATCAGCTTTACCTCCAGCTGGGAGAGATCAGTCCGTCTG 14546  
QY 312 cgttcggtggtggtcgtgaatgtacttt 339  
|||||  
DB 14545 CGTTTCGGTGGTGTCTGATGTTACTGT 14518

## RESULT 4

AAS37585  
ID AAS37585 standard; cDNA; 338 BP.AAS37585;  
AC AAS37585;  
XX  
XX 17-DEC-2001 (first entry)  
DT  
XX  
XX Novel human diagnostic and therapeutic gene #643.  
DE  
XX  
XX Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200166753-AZ.  
PN  
XX  
XX 13-SEP-2001.  
PD  
XX  
XX 09-MAR-2001; 2001WO-US07787.  
PE  
XX  
XX 09-MAR-2000; 2000US-0188609.  
PR  
XX  
XX(CHR) CHIRON CORP.  
(HYSE-) HYSEO INC.  
PA  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;  
PI Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leskowitz D, Kita D, Garcia V, Jones W, Stache-Crain B;  
PI  
XX  
XX WPI: 2001-530177/58.  
DR  
XX  
XX New polynucleotides and polypeptides, useful for diagnosis and  
PT treatment of breast, lung and colon cancer -  
PT  
XXClaim 1: page 757; 1193pp; English.  
XX  
XX The invention relates to new polynucleotides and polypeptides, useful for  
CC diagnosis and treatment of breast, lung and colon cancer. The sequences  
CC can be used in detecting differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample derived from a  
CC cell suspected of being cancerous. They can also be used to inhibit  
CC tumor growth by modulating expression of a gene product. AAS36943-  
CC AAS39338 represent novel human diagnostic and therapeutic coding  
CC sequences of the invention.  
CC  
XXSequence 338 BP; 93 A; 69 C; 91 G; 85 T; 0 other;  
SQ

Query Match 36.3%; Score 277.4; DB 22; Length 338;  
Best Local Similarity 98.2%; Pred. No. 2.9e-77;  
Matches 278; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 55 acaggttctgttllactctgaaaglaaactgagaactatatactgatactactact 114  
DB 56 agaggttctgttllactctgaaaglaaactgagaactatatactgatactactact 115  
|||||



Query Match	29.28;	Score 223.6;	DB 22,	Length 359,
-------------	--------	--------------	--------	-------------

CC antibodies against them can be used in pharmaceutical compositions to









FT	Key	Location/Qualifiers
FT	CDS	135...2741
FT		/tag=a
FT	sig_peptide	135..191
FT		/tag=b
FT	mat_peptide	192...2738
FT		/tag=c
PN		W09721811-A2.
XX		
PD		19-JUN-1997.
XX		
PF		13-DEC-1996; 96WU-US20696.
XX		
PR	10-MAY-1996;	96US-U044271.
PR	15-DEC-1995;	95US-0008657.
XX		
PA	(REG-)	REGENERON PHARM INC.
XX		
PI	Bowen DC, Glass DJ, Valenzuela DM, Yancopoulos GD;	
DR		WPI: 1997-332783/30.
XX		P-PSDB: AAM26610.
PT		
PT		
XX		
PS	Example 1; Fig 1; 120pp; English.	
CC		
CC	This cDNA sequence, deposited as ATCC 76498, includes a coding	
CC	sequence for a novel rat receptor tyrosine kinase designated muscle	
CC	specific kinase or MUSK (see AAM26610) that is expressed in normal	
CC	and denervated muscle. MUSK is alternatively referred to DMK for	
CC	denervated muscle kinase. The cDNA clone was isolated from	
CC	denervated muscle cDNA using primers (see AAT90476 and AAT90480)	
CC	based on tyrosine kinase homology domains. Human MUSK (see	
CC	AAT90473) has also been identified. Use of MUSK to generate anti-	
CC	MUSK antibodies and in the diagnosis of neurological or other	
CC	disorders is disclosed. Assay systems that may be used to detect	
CC	and/or measure ligands that bind the musk gene product are provided.	
CC	A claimed method of promoting the growth, differentiation or	
CC	survival of MUSK receptor-expressing cells involves administration	
CC	to the cell of agnif (see AAM26603). Such cells include muscle,	
CC	heart, spleen, ovary and retina cells, or cells genetically	
CC	engineered to express the MUSK receptor.	
CC		
XX		
SO	Sequence 2869 BF; 727 A; 754 C; 726 G; 658 T; 4 other;	
	Query Match	5.0%; Score 38.6; DB 18; Length 2869;
	Best Local Similarity	55.3%; Pred. NO. 0.22;
	Matches 99; Conservative 0; Mismatches 74; Indels 6; Gaps	1
QY	391 gtaagttggttgatggaatggaagccacccccaggttgaatgattgtgtacaaaac	450
DB		
DB	267 ggcagcttcctggtggccgttggaatccctccagccttgaaattcttggaccagaa	326
QY	451 agtagtctctcgatttagaagaagaacgcgtccacccaatccacagacaatgtgactllt	510
DB		
DB	327 aaattctccatcaagctgtttgaccccccgtcacagcatcgaggaacgycag	380
QY	511 caagtgcacatcccaagaatgtagaagactgtagcaacggaactcagttgatgtgcaaa	569
DB		
DB	381 ctctccaccctcctgagtgtaggagacagtgatgtgcatctactcgtcagacgcca	439
	RESULT 15	
	AAT87074	
	ID AAT87074 standard: cDNA: 2869 BP.	
XX		
XX	AAT87074:	
XX		

DT	06-JAN-1998	(first entry)
DE	Rat Dmk receptor cDNA.	
XX		
XX		
KW	dmk receptor; tyrosine kinase receptor; signal transduction; assay; therapy; diagnosis; ss.	
XX		
OS	Rattus sp.	
XX		
EH	Key	Location/Qualifiers
FT	CDS	135..2741
FT		/*tag= a
FT	sig_peptide	135..191
FT		/*tag= b
FT	mat_peptide	192..2738
FT		/*tag= c
PN	U55656473-A.	
XX		
PD	12-AUG-1997.	
XX		
PF	21-JUL-1993;	93US-0006558.
XX		
PR	19-JAN-1995;	95US-0374834.
PF	21-JUL-1993;	93US-0095658.
XX		
PA	(REGG-) PEGFNFPON PHARM INC.	
XX		
PI	Rojas EA, Valenzuela DM;	
XX		
TP	WPI: 1997-414593/38.	
DR	P-PSDB; AAM26507.	
XX		
PT	New isolated human denervated muscle kinase receptor - used to develop products for the diagnosis and treatment of neurological, muscle or neuromuscular disorders	
PT		
XX		
PS	Example 1; Column 23-28; 31pp; English.	
XX		
CC	This cDNA clone, deposited as ATCC 75498, codes for rat Dmk receptor (AAM26507), a novel tyrosine kinase receptor that is expressed at high levels in denervated muscle. It was isolated from a rat denervated skeletal muscle cDNA library following isolation of a partial sequence using primers (see AAT87075 and AAT87081) based on receptor tyrosine kinase (RTK) homology domains. PCR primers (AAT87075-81) based on homology domains between novel Dmk and RTKs can be used to identify novel RTKs. A human clone (see AAT87073) coding for a claimed human Dmk receptor (AAM26506) has been obtained.	
XX		
SO	Sequence 2869 BP; 727 A; 754 C; 726 G; 658 T; 4 other;	

Search completed. September 17, 2002, 10.01.53  
Job time: 6020 sec

.....

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 10:04:48 ; Search time 56.43 Seconds

(without alignments)  
432.513 Million cell updates/sec

Title: US-09-863-823-2

Perfect score: 1295  
Sequence: 1 MQGSPFLIVLFLPEPMYS RKKIMKCMKNCPISETAL 254

Scoring table: BLAST2M62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71:\*

1: pir1:\*\n2: pir2:\*\n3: pir3:\*\n4: pir4:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176	13.6	853	1 LIBONC	neural cell adhesi
2	175	13.5	858	1 LIBONC	neural cell adhesi
3	173.5	13.4	2783	2 T43416	hypothetical prote
4	171.5	13.2	946	1 A47299	ret-related recept
5	170	13.1	1896	2 T08851	Down syndrome cell
6	166	12.8	761	1 ICHUNG	neural cell adhesi
7	161	12.4	1011	2 T13669	neural cell adhesi
8	160.5	12.4	1091	1 IUCHNL	neural cell adhesi
9	159	12.3	725	1 ICHUNG	neural cell adhesi
10	159	12.3	1115	1 ICHUNG	neural cell adhesi
11	156	12.0	1173	2 T25893	hypothetical prote
12	153.5	11.9	1344	2 T14316	hypothetical prote
13	153	11.8	2029	1 TIEFELK	protein tyrosine p
14	152.5	11.8	5175	2 T29992	hypothetical prote
15	152.5	11.8	5198	2 T42290	hemocytin precurs
16	149	11.5	1447	2 A54100	tumor suppressor p
17	149	11.5	1912	2 A54178	protein-tyrosine p
18	148.5	11.5	338	2 JC4756	libic-system asso
19	148.5	11.5	4391	2 A38096	plecan precursor
20	146	11.3	7962	2 T38346	elastic titin ka
21	145.5	11.2	773	2 T46283	hypothetical prote
22	144.5	11.2	338	2 T46283	50k glycoprotein p
23	144.5	11.2	1348	2 S51656	vascular endotheli
24	144	11.1	458	1 WMSR1	biliary glycoprote
25	144	11.1	521	2 JC1508	biliary glycoprote
26	142.5	11.0	458	2 JC1509	biliary glycoprote
27	142.5	11.0	521	2 S34338	biliary glycoprote
28	142	11.0	871	1 T48696	protein-tyrosine k
29	142	11.0	881	1 T48697	protein-tyrosine k

30	142	11.0	1894	2 C54689	protein-tyrosine-p
31	141	10.9	976	2 T29583	hypothetical prote
32	139	10.7	1455	2 A44027	165k myofibrillar
33	138.5	10.7	797	2 S38579	fibroblast growth
34	138	10.7	1880	2 T18531	tractin - medictin
35	137.5	10.6	1898	2 S46216	leukocyte antigen
36	137	10.6	725	2 J60100	neural cell adhesi
37	137	10.6	1092	1 JN0635	neural cell adhesi
38	136.5	10.5	458	2 S23969	cell-adhesion mole
39	136.5	10.5	458	2 S68177	ecto-ATPase precu
40	136.5	10.5	519	2 A44783	leukocyte antigen
41	136.5	10.5	1897	1 T2405	sax-3 protein - Ca
42	135	10.4	1273	2 T14302	protein-tyrosine k
43	135	10.4	1356	2 JC4593	protein-tyrosine k
44	134.5	10.4	1070	2 T13822	fractioned gene prot
45	134	10.3	1375	2	

## ALIGNMENTS

RESULT 1  
LIBONC  
neural cell adhesion molecule short domain form precursor - bovine

N:Alternate names: NCAM-140  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 31-Mar-1993 #sequence:revision 31-Mar-1993 #text-change 22-Jun-1999  
C:Accession: A32976; A38778; B44290; S05402  
R:Lipkin, V.M.; Khrantsov, N.V.; Andreyeva, S.G.; Moslyakov, M.V.; Fetukhova, G.V.; R  
FENS lett. 254, 69-73, 1989

A>Title: Calmodulin-independent bovine brain adenylate cyclase. Amino acid sequence a  
A:Reference number: A32976; MUID:89378239  
A:Accession: A32976

A:Molecule type: mRNA  
A:Residues: 1-853 <IDP>  
A:Cross-references: GB:X16451; MID:960; PIDN:CAA34470.1; PID:961

A:Accession: A38778  
A:Molecule type: protein  
A:Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-5  
A:Note: The authors identified this protein as calmodulin-independent adenylate cycl  
R:Rougon, G.; Marshak, D.R.  
J. Biol. Chem. 261, 3396-3401, 1986

A>Title: Structural and immunological characterization of the amino-terminal domain o  
A:Reference number: A44290; MUID:86140120  
A:Accession: B44290

A:Molecule type: protein  
A:Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-5  
A:Note: 23-31u was also found  
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM  
C:Comment: Various forms of NCAM are produced by alternative splicing.  
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; I  
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;  
F:1-13/Domain: signal sequence #status predicted <SIG>  
F:20-853/Product: neural cell adhesion molecule, short domain form #status experiment  
E:20-719/Domain: extracellular; #status predicted <EXT>  
E:34-38/Domain: immunoglobulin; #status predicted <IM1>  
E:132-191/Domain: immunoglobulin homology; #IM2>  
F:122-156/Region: heparin binding #status predicted  
F:161-165/Region: heparin binding #status predicted  
F:228-288/Domain: immunoglobulin homology <IMM3>  
F:261-270/Region: NCAM binding #status predicted  
F:321-396/Domain: immunoglobulin homology <IMM4>  
F:428-490/Domain: immunoglobulin homology <IMM5>  
E:527-604/Domain: fibronectin type III repeat homology; #FN3A>  
F:633-693/Domain: fibronectin type III repeat homology; #FN3B>  
F:720-737/Domain: transmembrane #status predicted <TM>  
F:738-853/Domain: intracellular #status predicted <INT>  
F:41-95;159-169;235-286;326-394;435-488/Disulfide bonds: #status predicted  
F:222;314;346;432;458;487/Binding site: carbohydrate (Asn) (covalent); #status predict

Query Match 13.6%; Score 176; DB 1; Length 853;  
Best Local Similarity 28.4%; Pred No 7 7e-06;

Matches: 60; Conservative: 31; Mismatches: 88; Indels: 32; Gaps: 8;

QY 41 QASLIIC-----AVGNHREELLMYREKGVLDKSNKINS-----SSVCS 82

Db 423 QVITTCASDPIPTSTWRTSTNISEEKASMTREKQETLLDGMVVSMAKVSSTLIK 382

QY 83 SIFPNMSTFTC-----RIKIDQSVSVAVIANTPEPPIISGNFQTVESNVCACGNK 138

Db 483 SIQYTDAG-EYCTASNTICOD---SOSMYLEVOYAPKIQGPVAVYTWGNOVNITCEVF 438

QY 139 ANPQAMMMYKNS-S-LIIEP-SHQVLSSTVUSLTKVFP+QNGTYSCLASSLIKTES 197

Db 439 AYPASATISWEPDGLDPSNSYNLKIYNTPSASYLEVTPSEDFGNMVTANPIGDES 498

QY 198 LDFHLLVWKDKVGVF-----LEPTAACVVF 224

Db 499 LEPVLYQAD-TPSSPSTIDVPEPSSIAQVVF 528

RESULT 2

Query Match 13.58; Score 175; DB 1; Length 856;  
best local similarity 28.48; pred. no. 9, 30-66;  
Matches: 60; Conservative: 31; Mismatches: 88; Indels: 32; Gaps: 8;

QY 41 QASLIIC-----AVGNHREELLMYREKGVLDKSNKINS-----SSVCS 82

Db 423 QVITTCASDPIPTSTWRTSTNISEEKASMTREKQETLLDGMVVSMAKVSSTLIK 382

QY 83 SIFPNMSTFTC-----RIKIDQSVSVAVIANTPEPPIISGNFQTVESNVCACGNK 138

Db 483 SIQYTDAG-EYCTASNTICOD---SOSMYLEVOYAPKIQGPVAVYTWGNOVNITCEVF 438

QY 139 ANPQAMMMYKNS-S-LIIEP-SHQVLSSTVUSLTKVFP+QNGTYSCLASSLIKTES 197

Db 439 AYPASATISWEPDGLDPSNSYNLKIYNTPSASYLEVTPSEDFGNMVTANPIGDES 498

QY 198 LDFHLLVWKDKVGVF-----LEPTAACVVF 224

Db 499 LEPVLYQAD-TPSSPSTIDVPEPSSIAQVVF 528

RESULT 3

Query Match 13.48; Score 171.5; DB 2; Length 2783;  
best local similarity 23.18; pred. no. 7, 65;  
Matches: 60; Conservative: 36; Mismatches: 100; Indels: 53; Gaps: 9;

QY 41 QASLIIC-----AVGNHREELLMYREKGVLDKSNKINS-----SSVCS 82

Db 423 QVITTCASDPIPTSTWRTSTNISEEKASMTREKQETLLDGMVVSMAKVSSTLIK 382

QY 83 SIFPNMSTFTC-----RIKIDQSVSVAVIANTPEPPIISGNFQTVESNVCACGNK 138

Db 483 SIQYTDAG-EYCTASNTICOD---SOSMYLEVOYAPKIQGPVAVYTWGNOVNITCEVF 438

QY 139 ANPQAMMMYKNS-S-LIIEP-SHQVLSSTVUSLTKVFP+QNGTYSCLASSLIKTES 197

Db 439 AYPASATISWEPDGLDPSNSYNLKIYNTPSASYLEVTPSEDFGNMVTANPIGDES 498

QY 198 LDFHLLVWKDKVGVF-----LEPTAACVVF 224

Db 499 LEPVLYQAD-TPSSPSTIDVPEPSSIAQVVF 528

RESULT 4

Query Match 13.48; Score 171.5; DB 2; Length 2783;  
best local similarity 23.18; pred. no. 7, 65;  
Matches: 60; Conservative: 36; Mismatches: 100; Indels: 53; Gaps: 9;

QY 41 QASLIIC-----AVGNHREELLMYREKGVLDKSNKINS-----SSVCS 82

Db 423 QVITTCASDPIPTSTWRTSTNISEEKASMTREKQETLLDGMVVSMAKVSSTLIK 382

QY 83 SIFPNMSTFTC-----RIKIDQSVSVAVIANTPEPPIISGNFQTVESNVCACGNK 138

Db 483 SIQYTDAG-EYCTASNTICOD---SOSMYLEVOYAPKIQGPVAVYTWGNOVNITCEVF 438

QY 139 ANPQAMMMYKNS-S-LIIEP-SHQVLSSTVUSLTKVFP+QNGTYSCLASSLIKTES 197

Db 439 AYPASATISWEPDGLDPSNSYNLKIYNTPSASYLEVTPSEDFGNMVTANPIGDES 498

QY 198 LDFHLLVWKDKVGVF-----LEPTAACVVF 224

Db 499 LEPVLYQAD-TPSSPSTIDVPEPSSIAQVVF 528

Matches: 60; Conservative: 31; Mismatches: 88; Indels: 32; Gaps: 8;

QY 41 QASLIIC-----AVGNHREELLMYREKGVLDKSNKINS-----SSVCS 82

Db 423 QVITTCASDPIPTSTWRTSTNISEEKASMTREKQETLLDGMVVSMAKVSSTLIK 382

QY 83 SIFPNMSTFTC-----RIKIDQSVSVAVIANTPEPPIISGNFQTVESNVCACGNK 138

Db 483 SIQYTDAG-EYCTASNTICOD---SOSMYLEVOYAPKIQGPVAVYTWGNOVNITCEVF 438

QY 139 ANPQAMMMYKNS-S-LIIEP-SHQVLSSTVUSLTKVFP+QNGTYSCLASSLIKTES 197

Db 439 AYPASATISWEPDGLDPSNSYNLKIYNTPSASYLEVTPSEDFGNMVTANPIGDES 498

QY 198 LDFHLLVWKDKVGVF-----LEPTAACVVF 224

Db 499 LEPVLYQAD-TPSSPSTIDVPEPSSIAQVVF 528

RESULT 3

Query Match 13.48; Score 171.5; DB 2; Length 2783;  
best local similarity 23.18; pred. no. 7, 65;  
Matches: 60; Conservative: 36; Mismatches: 100; Indels: 53; Gaps: 9;

QY 41 QASLIIC-----AVGNHREELLMYREKGVLDKSNKINS-----SSVCS 82

Db 423 QVITTCASDPIPTSTWRTSTNISEEKASMTREKQETLLDGMVVSMAKVSSTLIK 382

QY 83 SIFPNMSTFTC-----RIKIDQSVSVAVIANTPEPPIISGNFQTVESNVCACGNK 138

Db 483 SIQYTDAG-EYCTASNTICOD---SOSMYLEVOYAPKIQGPVAVYTWGNOVNITCEVF 438

QY 139 ANPQAMMMYKNS-S-LIIEP-SHQVLSSTVUSLTKVFP+QNGTYSCLASSLIKTES 197

Db 439 AYPASATISWEPDGLDPSNSYNLKIYNTPSASYLEVTPSEDFGNMVTANPIGDES 498

QY 198 LDFHLLVWKDKVGVF-----LEPTAACVVF 224

Db 499 LEPVLYQAD-TPSSPSTIDVPEPSSIAQVVF 528

RESULT 4

Query Match 13.48; Score 171.5; DB 2; Length 2783;  
best local similarity 23.18; pred. no. 7, 65;  
Matches: 60; Conservative: 36; Mismatches: 100; Indels: 53; Gaps: 9;

QY 41 QASLIIC-----AVGNHREELLMYREKGVLDKSNKINS-----SSVCS 82

Db 423 QVITTCASDPIPTSTWRTSTNISEEKASMTREKQETLLDGMVVSMAKVSSTLIK 382

QY 83 SIFPNMSTFTC-----RIKIDQSVSVAVIANTPEPPIISGNFQTVESNVCACGNK 138

Db 483 SIQYTDAG-EYCTASNTICOD---SOSMYLEVOYAPKIQGPVAVYTWGNOVNITCEVF 438

QY 139 ANPQAMMMYKNS-S-LIIEP-SHQVLSSTVUSLTKVFP+QNGTYSCLASSLIKTES 197

Db 439 AYPASATISWEPDGLDPSNSYNLKIYNTPSASYLEVTPSEDFGNMVTANPIGDES 498

QY 198 LDFHLLVWKDKVGVF-----LEPTAACVVF 224

Db 499 LEPVLYQAD-TPSSPSTIDVPEPSSIAQVVF 528

for-related receptor RTK - Pacific electric ray  
M:Contains: protein tyrosine kinase (EC 2.7.1.112)  
C:Species: Torpedo californica (Pacific electric ray)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-May-2000  
C:Accession: A47299  
R:Jennings, C.G.; Dyer, S.M.; Burden, S.J.  
Proc. Natl. Acad. Sci. U.S.A. 90, 2895-2899, 1993  
A:Title: Muscle-specific trk-related receptor with a kinase domain defines a distinct C  
A:Reference number: A47299; MUID:93219391  
A:Accession: A47299  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-946 <JFN>  
A:Cross-references: GB:U11311; NID:q290857; PUDN:AAA4285.1; PID:q290858  
A:Experimental source: electric organ  
A:Note: Sequence extracted from NCBI backbone (NCBI:128724, NCBI:128726)  
C:Superfamily: Torpedo for-related receptor; immunoglobulin homology; kinase homology;  
C:Keywords: ATP, glycoprotein, kinase, phosphotransferase; transmembrane protein; tyros  
F:44-103/Domain: immunoglobulin homology <IM1>  
F:137-195/Domain: immunoglobulin homology <IM2>  
F:229-287/Domain: immunoglobulin homology <IM3>  
F:464-542/Domain: kinase homology <KR3>  
F:572-588/Domain: transmembrane #status predicted <TM>  
F:650-940/Domain: protein kinase homology <KIN>  
F:658-666/Region: protein kinase ATP-binding motif  
F:225,340,477,544/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 13.28; Score 171.5; DB 1; Length 946;  
Best Local Similarity 25.18; Pred. No. 20-05;  
Matches 63; Conservative 38; Mismatches 95; Indels 55; Gaps 10;

7 LLVLVLT-----FLPR-EMTSVITNGKTEYILDTTSGOASLICAVONTREELMYREGRVLDKGNK- 73  
10 LLVLTGSSADGILPKAPQITPLEYDALE-----EFASFCVADSYV-AA 59  
56 ELLWYREGRV---DLKSGNKINSSVCSSISENDNGI-SFTGRLGRDQSVSVYLVN 111  
60 EITWTRNNIPRPDTRKSTKENCQILITLSEDTDNGVCCTTANNMGSSASCGALQY 119  
112 TFP--LISGNDPQTVRESGNVKLVCKYKANKPAOMMYKNSLLEKSRHOIQOTSES 169  
120 KMKPKTIRPPDGVKAL-LGSKVYVPCSMGNPKPAISWFKDETALKNQDRTSV---LES 175  
170 FGLSTIKYKPEKNGYSCIAKSLKTE-----SILDFHLIVK 205  
176 GNLRIKNVQLEDAKGYRCLARRNSLGEYSRGALEVOVSARIVAKPTSQNVSYGSEVILQ 235  
205 DRTVGVPTFPI 216  
236 CKATGFPIPTI 246

RESULT 5  
T08851  
Down syndrome cell adhesion protein 1 - human (fragment)  
N:Alternate names: Down syndrome cell adhesion molecule  
C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
C:Accession: T08851  
R:Yamakawa, K.; Huo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.; Kutenbetz  
submitted to the EMBL Data Library, September 1997  
A:Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down sy  
A:Reference number: Z16495  
A:Accession: T08851  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1896 <YAM>  
A:Cross-references: EMBL:AF023449; NID:q3169765; PID:q3169766  
A:Experimental source: brain; developmental stage: 14 weeks; fetal  
C:Genetics:  
A:Gene: DSCAM  
A:Map position: 21q22

A:Note: derived from alternatively-spliced mRNA  
C:Function:  
A:Description: involved in nervous system development  
C:Keywords: alternative splicing

Query Match 13.18; Score 170; DB 2; Length 1896;  
Best Local Similarity 29.88; Pred. No. 5.9e-05;  
Matches 57; Conservative 31; Mismatches 69; Indels 34; Gaps 10;

15 PREMTSVITNGKTEYILDTTSGOASLICAVONTREELMYREGRVLDKGNK- 73  
306 PRKXSSV-----GSQVSLSCSTG-TEQELSWR-NEEL-LNGKNV 346  
74 ----INSSVCSSISENDNGISFTGRLGRDQSV--SVVLNVTFPPLISGNDPQTV 126  
347 RITGTHNHLIMDHYVSDG-AVQCFVRKDISAQDYVGVLEDTPKIISAFSEKVV 405  
127 ESNVNLVCKYKANKPAOMMYKNSLLEKSRHOIQOTSES---FGLSTIKYKPE 181  
406 PAEPFSLCKNVKGTPLFTITWLTLEDT LKGSRLSCMITSEGNVSYLNISSQVAD 464  
182 NQVSYCAKSS 192  
465 GGVYRCTANN 475

RESULT 6  
10HUNG  
neural cell adhesion molecule 1 gp1-anchored splice form precursor, muscle-specific -

N:Alternate names: CD56; NCAM-120  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 28-Jan-2000  
C:Accession: S07784; A26883  
R:Barton, C.H.; Dickson, G.; Gower, H.J.; Rowett, L.H.; Putt, W.; Elsom, V.; Moore, S  
Development 104, 165-173, 1988  
A:Title: Complete sequence and in vitro expression of a tissue-specific phosphatidyli  
A:Reference number: S07784; MUID:89305258  
A:Accession: S07784  
A:Molecule type: mRNA  
A:Residues: 1-761 <BAR>  
A:Cross-references: EMBL:X16841; NID:q35005; PUDN:CAA34739.1; PID:q35006  
R:Dickson, G.; Gower, H.J.; Barton, C.H.; Prentice, H.M.; Elsom, V.L.; Moore, S.E.; C  
Cell 50, 1119-1130, 1987  
A:Title: Human muscle neural cell adhesion molecule (N-CAM): identification of a musc  
A:Reference number: A90895; MUID:87301755  
A:Accession: A26883  
A:Molecule type: mRNA  
A:Residues: 491-761 <DJC>  
A:Cross-references: GB:M17409; NID:q189097; PUDN:AAA59912.1; PID:q386979  
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM  
C:Comment: Various forms of NCAM are produced by alternative splicing.  
C:Genetics:  
A:Gene: GDB:NCAM1; NCAM; CD56  
A:Cross-references: GDB:119448; OMIM:116930  
A:Map position: 11q23.2, 11q23.3  
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology: 1  
C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membra  
F:1-19/Domain: signal sequence #status predicted <Sig>  
F:20-761/Product: neural cell adhesion molecule phosphatidylinositol-linked form, mus  
F:34-98/Domain: immunoglobulin homology <IM1>  
F:132-191/Domain: immunoglobulin homology <IM2>  
F:152-166/Region: heparin binding #status predicted  
F:161-165/Region: heparin binding #status predicted  
F:228-287/Domain: immunoglobulin homology <IM3>  
F:263-272/Region: NCAM binding #status predicted  
F:322-387/Domain: immunoglobulin homology <IM4>  
F:419-481/Domain: immunoglobulin homology <IM5>  
F:499-587/Domain: fibronectin type III repeat homology #status atypical <FN3A>  
F:633-720/Domain: fibronectin type III repeat homology #status atypical <FN3B>  
F:41-96, 139-189, 235-287, 329-385, 426-479/Disulfide bonds: #status predicted  
F:222, 315, 347, 423, 449, 478/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 12.4% Score 161; 28 2; Length 1011;  
 best local similarity 24.1% from May 9 1997;  
 Matches 46; Conservative 34; Mismatches 81; Indels 28; Gaps 6;  
 QY 23 LTVNCKTENVLTIDTFGQASLTAVQNHTRKEELTWY-----EEGRVDL----- 68  
 DB 245 IDINQVNI FVQGVAVI DTDG APTAVN IWNITITL IISGND IIVVSWET 370  
 QY 69 KSNQKINSSVSVSSISBNQSLST-----KAGGKSVSVVLNLTHTLSSGNO 121  
 DB 421 KSDITHTQSLTINATPEEIMVPEEFAFNIVITLNERKTSALTEVYPPVAVVSP 350  
 QY 122 FGVTEGQNVKLVGVNAPVLA TGMNMTNSSLSSGNSGNGVGTSGVSLVPE 178  
 DB 981 SALTANSEVLNLTETVAPVASTLVGVMTKNDLIYVNDITTHKGSSEVALYKSTE 440  
 QY 179 KPIQRTVSG 187  
 DB 441 KEDIGNYSG 449  
 RESULT 8  
 FUNCTION  
 neutral cell adhesion molecule, long domain form precursor - chicken  
 NCBI alternate names: NCAM 180  
 NCBI aliases: neutral cell adhesion molecule, short domain form (NCAM-140)  
 CSpecies: Gallus gallus (chicken)  
 CDate: 41; Mar 1993 #sequence=15; 31; Mar 1993 #blast change 22; Jun 1993  
 CAccession: A43613; B43613; A254335; B254335; A46550; S36950; A44369; A60852; S29668  
 CContributor: R.A.; Hemperly, J.; Moriya, R.A.; Friediger, E.A.; Brackenbury, R.; Edlin  
 CRef: Science 270, 799-806, 1987

A:Title: Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell  
A:Reference number: A43613; MUID:87206190  
A:Accession: A43613  
A:Molecule type: mRNA  
A:Residues: 1-175 <CU2>  
A:Cross-references: GB:M15860  
A:Accession: A43613  
A:Molecule type: Protein  
A:Residues: 26-44,120-127,292-321,320-342,393-415,640-650,822-828 <QUN>  
A:Mol. Asp. 222 probably binds carbohydrate, Asp 226 probably does not  
R:Hampey, J.J.; Murray, B.A.; Edelman, G.M.; Cunningham, B.A.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3037-3041, 1986  
A:Title: sequence of a cDNA clone encoding the polysialic acid-rich and cytoplasmic d  
A:Reference number: A25435; MUID:86208089  
A:Accession: A25435  
A:Molecule type: mRNA  
A:Residues: 128-1091 <HEM>  
A:Cross-references: GB:M13210  
A:Accession: B25435  
A:Molecule type: Protein  
A:Residues: 128-166,472-480,428-439,611-631,744-769,765-781,1040-1084 <HE>  
R:Murray, B.A.; Owens, G.C.; Prediger, E.A.; Crossin, K.L.; Cunningham, B.A.; Edelman  
J. Cell Biol. 103, 1411-1419, 1986  
A:Title: Cell surface modulation of the neural cell adhesion molecule resulting from  
A:Reference number: A46550; MUID:87033934  
A:Accession: A46550  
A:Molecule type: DNA  
A:Residues: 810-1070 <MUR>  
A:Cross-references: GB:X04479  
R:Sahner, M.; Covault, J.  
submitted to the EMBL data library, February 1993  
A:Reference number: S69950  
A:Accession: S69950  
A:Molecule type: DNA  
A:Residues: 1-117 <SAS>  
A:Cross-references: EMBL:X70342; MUID:9417643; FID:CA449807 1; FID:9417632  
R:Wells, G.L.; H. Forrest, D.; Brackenbury, R.  
Genomics 14, 875-882, 1992  
A:Title: Conserved regulatory elements in the promoter region of the N-CAM gene.  
A:Reference number: A4369; MUID:31122797  
A:Accession: A4369  
A:Molecule type: DNA  
A:Residues: 1-117 <GOS>  
A:Cross-references: FMR:J12128; MUID:363653; FID:CA448113 1; FID:363654  
A:Experimental source: White Leghorn  
R:Cole, G.J.; Leewy, A.; Cross, N.V.; Akesson, R.; Glaser, L.  
J. Cell Biol. 103, 1739-1744, 1986  
A:Title: Topographic localization of the heparin-binding domain of the neural cell ad  
A:Reference number: A4652; MUID:87057627  
A:Accession: A60852  
A:Molecule type: Protein  
A:Residues: 20-29 <GOS>  
R:van, Y.; Wu, X.F.; Gately, J.; Rothbauer, U.; Stu, C.H.  
J. Cell Biol. 118, 947-949, 1992  
A:Title: Identification of a peptide sequence involved in homophilic binding in the n  
A:Reference number: A43280; MUID:92363934  
A:Contents: annotation: homophilic binding region  
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM  
C:Comment: Various forms of NCAM are produced by alternative splicing.  
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; 1  
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;  
E:1-19/Domain: signal sequence #status predicted <SIG>  
E:20-1691/Product: neural cell adhesion molecule, long domain form #status experiment  
E:20-809,1071-1091/Product: neural cell adhesion molecule, short domain form #status  
E:20-711/Domain: extracellular #status predicted <EXT>  
E:34-96/Domain: immunoglobulin homology <IMM1>  
E:132-131/Domain: immunoglobulin homology <IMM2>  
E:152-156/Region: heparin binding #status predicted  
E:161-165/Region: heparin binding #status predicted  
E:128-287/Domain: immunoglobulin homology <IMM3>  
E:262-271/Region: NCAM binding #status experimental  
E:322-367/Domain: immunoglobulin homology <IMM4>  
E:419-481/Domain: immunoglobulin homology <IMM5>



F:518-595/Domain: fibronectin type III repeat homology <FN3>  
 F:624-685/Domain: fibronectin type III repeat homology <FN3>  
 F:712-729/Domain: transmembrane #status predicted <TM>  
 F:730-1091/Domain: intracellular #status predicted <INT>  
 F:41-96,139-189,235-287,329-385,426-479/Disulfide bonds: #status predicted  
 F:222/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:315,347,423,449,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 12.4% Score 160.5; DB 1; Length 1091;  
 Best Local Similarity 26.9%; Pred. No. 0.00017;  
 Matches 60; Conservative 37; Mismatches 87; Indels 39; Gaps 10;

25 VNGKTEYILDTPGSSQASLC-----AVQNHTRREELLWREGGRVDELS 70  
 DB 313 VENKAMELED-----QITLCEASGDPISITWKTSTRNISNEKTL-----DGRIVRS 363  
 QY 71 GNKINSSVCVSSISENNNGISFTC-----RLGRDQSVSVLVNTPPILSGNDFOYVE 126  
 DB 364 HARV--SSLTKETIYIDAG-EYVCTASNTIGD---SQAMYLEVYAPKLGQPVAVVYTM 417  
 QY 127 EGSNVLKVCNKANPOAQMVMYKNSLLEDEK-SRHQIQOTSESFOLSTIKYKPNNGTY 185  
 DB 418 EGNQVNTICEVFAVPASAVISMFRRGQLPSSNYSNIKIYTPASASTYLEVTPDSENFQNT 477  
 QY 186 SCIAKSLKTSIDPHILVNDKTYGV-----IEPIAACVIF 224  
 DB 478 NCTAVNRIGDSSSEFTIQQAD-TPSSPSIDRVEPYSSTARVEF 519

## RESULT 9

IUMSG

neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse

N:Alternate names: NCAM-120

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text change 31-Dec-2000

C:Accession: A29673; S00382; A44290

R:Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.P.; Ponteg

EMBO J. 6, 907-914, 1987

A:Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000

A:Reference number: A29673; MUID:87246524

A:Accession: A29673

A:Molecule type: mRNA

A:Residues: 1-725 &lt;BAR&gt;

A:Cross-references: EMBL:Y00051; NID:953342; PIDN:CAA68263.1; PID:953343

R:Barthels, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.

EMBO J. 7, 625-632, 1988

A:Title: Differential splicing and alternative polyadenylation generates distinct NCAM

A:Reference number: S00382; MUID:88283628

A:Accession: S00382

A:Molecule type: DNA

A:Residues: 642-656, 'D', 658-725 &lt;BA2&gt;

A:Cross-references: EMBL:X07195

R:Rougon, G.; Marshak, D.R.

J. Biol. Chem. 261, 3396-3401, 1986

A:Title: Structural and immunoglobulin characterization of the amino-terminal domain of m

A:Reference number: A44290; MUID:86140120

A:Accession: A44290

A:Molecule type: protein

A:Residues: 20-36 &lt;ROU&gt;

C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol

C:Comment: Several forms of NCAM are produced by alternative splicing. See also RIR.IUMS

C:Genetics:

A:Gene: NCAM

A:Map position: 9

A:Introns: 701/1

C:Superfamily: neural cell adhesion molecule: fibronectin type III repeat homology; immu

C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane

F:228-230/Domain: immunoglobulin homology <IMM>  
 F:263-272/Region: NCAM binding #status predicted  
 F:323-388/Domain: immunoglobulin homology <IMM>  
 F:420-482/Domain: immunoglobulin homology <IMM>  
 F:519-526/Domain: fibronectin type III repeat homology <FN3>  
 F:625-685/Domain: fibronectin type III repeat homology <FN3>  
 F:41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted  
 F:222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 12.3% Score 159; DB 1; Length 725;  
 Best Local Similarity 27.5%; Pred. No. 0.00013;  
 Matches 57; Conservative 35; Mismatches 81; Indels 34; Gaps 9;

41 QASLIC-----AVQNHTRREELLWREGGRVDELSGNKINSSVCVSSISE 86  
 DB 325 QVTLCEASGDPISITWKTSTRNISSEEDL-----DGHVYVASHARV--SSLTKSLGY 378  
 QY 87 NDNGISFTC-----RLGRDQSVSVLVNTPPILSGNDFOYVEESSNKLVCNKANPO 142  
 DB 379 RDAG-EYVCTASNTIGD---SQSIDLEPQYAPKLGQPVAVVYMEGNQVNTICEVAPPS 434  
 QY 143 AQMMYKNSLLEDEK-SRHQIQOTSESFOLSTIKYKPNNGTYSCIAKSLKTSIDPH 201  
 DB 435 ATISMFRRGQLPSSNYSNIKIYTPASASTYLEVTPDSENFQNTCTAVNRIGDSELEFI 494  
 QY 202 LIVDKTYGVF-----IEPIAACVIF 224  
 DB 495 LVQAD-TPSSPSIDRVEPYSSTARVQVF 520

## RESULT 10

IUMSG

neural cell adhesion molecule 1 precursor, long domain splice form - mouse

N:Alternate names: NCAM-180

N:Contains: neural cell adhesion molecule, short domain splice form (NCAM-140)

C:Species: Mus musculus (house mouse)

C:Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text change 31-Dec-2000

C:Accession: A29673; S00844; S00384; A28281; A44290; S00383

R:Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Pon

EMBO J. 6, 907-914, 1987

A:Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,

A:Reference number: A29673; MUID:87246524

A:Accession: A29673

A:Molecule type: mRNA

A:Residues: 1-548, 'T', 550-571, 'T', 573-574, 'D', 576-588, 'WQPS', 593, 'S', 595-599, 'P', 601,

A:Cross-references: EMBL:Y00051; NID:953342; PIDN:CAA68263.1; PID:953343

R:Barthels, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.

Nucleic Acids Res. 15, 8621-8641, 1987

A:Title: Analysis of cDNA clones that code for the transmembrane forms of the mouse n

A:Reference number: S00844; MUID:88067687

A:Accession: S00844

A:Molecule type: mRNA

A:Residues: 529-809,1077-1115 &lt;SAN&gt;

A:Cross-references: EMBL:X06328; NID:953342; PIDN:CAA29641.1; PID:9817984

R:Barthels, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.

EMBO J. 7, 625-632, 1988

A:Title: Differential splicing and alternative polyadenylation generates distinct NCA

A:Reference number: S00382; MUID:86243628

A:Accession: S00384

A:Molecule type: DNA

A:Residues: 642-1115 &lt;BAR&gt;

A:Cross-references: EMBL:X07195

R:Barthels, D.; Vopper, G.; Wille, W.

Nucleic Acids Res. 16, 4217-4225, 1988

A:Title: NCAM-180, the large isoform of the neural cell adhesion molecule of the mous

A:Reference number: A28281

A:Accession: A28281

A:Molecule type: mRNA

A:Residues: 804-1081 &lt;BA3&gt;

A:Cross-references: EMBL:X07244; NID:953342; PIDN:CAA30230.1; PID:9529720

R:Rougon, G.; Marshak, D.R.

J. Biol. Chem. 261, 3396-3401, 1986

Accession: A44290  
 A:Accession: A44290; MIM:86140120  
 A:Reference: 20 46 - protein  
 A:Keywords: MAM mediates cell-cell adhesion via homophilic binding with another MAM and C-terminal: Several forms of MAM are produced by alternative splicing. See also ETK:JMS  
 A:Gene: NCAM  
 A:Map position: 9  
 A:Intron: 64/41 70/41 77/42 83/42 103/42  
 C:Superfamily: neural cell adhesion molecule, fibronectin type II: repeat homology, immunoglobulin, alternative splicing, brain, cell adhesion, duplication, heparin binding, 51  
 E:1-19/Domain: signal sequence #status predicted 812  
 E:20-111/Domain: neural cell adhesion molecule, long domain splice form #status expert  
 E:20-809/1077-1115/Product: neural cell adhesion molecule, short domain splice form #sta  
 E:20-711/Domain: extracellular #status predicted TEXT  
 E:34-98/Domain: immunoglobulin homology (IMM)  
 E:142-191/Domain: immunoglobulin homology (IMM)  
 E:152-156/Region: heparin binding #status predicted  
 E:161-165/Region: heparin binding #status predicted  
 E:228-299/Domain: immunoglobulin homology (IMM)  
 E:262-272/Region: NCAM binding #status predicted  
 E:323-388/Domain: immunoglobulin homology (IMM)  
 E:420-482/Domain: immunoglobulin homology (IMM)  
 E:519-596/Domain: fibronectin type III repeat homology (FN3)  
 E:625-682/Domain: fibronectin type III repeat homology (FN3)  
 E:712-729/Domain: transmembrane #status predicted TM  
 E:741-745/Domain: cytoplasmic #status predicted CM  
 E:841-961/9189-245288,330-380,427-480/Disulfide bonds: #status predicted  
 E:222-316,348-424,453-479/Binding sites: antibody (Ab) (occasionally #status predicted)

Query Match 12.48; Score 156; DB 2; Length 1173;  
 Best Local Similarity: 29.78; Field No. 0.00041;  
 Matches 57; Conservative 45; Mismatches 81; Indels 34; Gaps 9.

41 CASLIG AVQHTREELMTREHPVPLSGSKINSSVGVSSISF 96  
 127 QVTLTEASRQIPISITWRTSTRNISRQDIDHVVRSRAV--SILTIKSIQY 378  
 87 NINQISLPE---RLRGSVSVSVLVNIFITLISNLQIYELSSNVEVNVNANQ 432  
 100 479 KANG EYKATASNTIGQD---SGLDLPQYATKLGQVNAVYTWEGQVNIITEVFAIPS 434  
 143 AQMMYKNSSTIQK--SPGTEGTPSPQI STTKVKKPNQGYCTAKSKLTPSI DPH 201  
 145 435 ATLSM-RKQGLPSSNSYNTIKTYNTPASYLEVTPUSDENGNCTAVNRICDESLEFI 494  
 QY 292 LKVKRYGVV---LPIIACVPI 224  
 145 495 LVQAD-TTSSPSIDKVPYSSPAQVQF 520

RESULT 11  
 T25894  
 Hypothetical protein T1984.7 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15 Oct 1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T25894  
 R:Gallway, S.  
 submitted to the EMBL data library, November 1996  
 A:Description: The sequence of C. elegans cosmid T1984.  
 A:Reference number: 220106  
 A:Accession: T25894  
 A:Status: preliminary; translated from cDNA/EMBL/TrEMBL  
 A:Molecule type: DNA  
 A:Residues: 1-1173 (CDS)  
 A:Cross-references: EMBL:U08438; PDB:AA3765.1; GSPDB:GN00019; CDS: T1984.7  
 A:Experimental source strain Bristol N2; clone T1984  
 C:Genetics:  
 A:Gene: CDS: T1984.7  
 A:Map position: 1

Accession: A46182  
 A:Accession: A46182; MIM:90046860  
 A:Reference: 20 46 - protein  
 A:Keywords: MAM mediates cell-cell adhesion via homophilic binding with another MAM and C-terminal: Several forms of MAM are produced by alternative splicing. See also ETK:JMS  
 A:Gene: NCAM  
 A:Map position: 9  
 A:Intron: 64/41 70/41 77/42 83/42 103/42  
 C:Superfamily: neural cell adhesion molecule, fibronectin type II: repeat homology, immunoglobulin, alternative splicing, brain, cell adhesion, duplication, heparin binding, 51  
 E:1-19/Domain: signal sequence #status predicted 812  
 E:20-111/Domain: neural cell adhesion molecule, long domain splice form #status expert  
 E:20-809/1077-1115/Product: neural cell adhesion molecule, short domain splice form #sta  
 E:20-711/Domain: extracellular #status predicted TEXT  
 E:34-98/Domain: immunoglobulin homology (IMM)  
 E:142-191/Domain: immunoglobulin homology (IMM)  
 E:152-156/Region: heparin binding #status predicted  
 E:161-165/Region: heparin binding #status predicted  
 E:228-299/Domain: immunoglobulin homology (IMM)  
 E:262-272/Region: NCAM binding #status predicted  
 E:323-388/Domain: immunoglobulin homology (IMM)  
 E:420-482/Domain: immunoglobulin homology (IMM)  
 E:519-596/Domain: fibronectin type III repeat homology (FN3)  
 E:625-682/Domain: fibronectin type III repeat homology (FN3)  
 E:712-729/Domain: transmembrane #status predicted TM  
 E:741-745/Domain: cytoplasmic #status predicted CM  
 E:841-961/9189-245288,330-380,427-480/Disulfide bonds: #status predicted  
 E:222-316,348-424,453-479/Binding sites: antibody (Ab) (occasionally #status predicted)

Query Match 12.08; Score 156; DB 2; Length 1173;  
 Best Local Similarity: 26.48; Field No. 0.00041;  
 Matches 57; Conservative 38; Mismatches 85; Indels 36; Gaps 10.

7 LILVILFLPREMTSSVLTNGKTEYILDTTP-----GSSGLICAVQNTREBELM 59  
 191 IIAKVTIKGRPFSEIV-----SNEIVFITPQVAVKADHLEELVASTLR--QYK 244  
 60 YEEERK--VQLSKNKNSSVSVSSISENDNISFTGL--GADQSVSVYVNTFFPL 116  
 245 TRSPQIIVDCVPIRPVYSSSIIVSPASITFTGL--VTPASNNNSIPAVSVYVAPP 303  
 117 LSGNPFQVE--PSSNVKIVC--NVKANPQAMMYKN-----SLIDLEKSRHQLQUTSE 168  
 404 ITTPRTKVAATFADVLEFCTAARPAEAVNNYKNEALIGSEYVIEPK----- 355  
 169 SPQISTKVERKPNQGYCTAKSKLKTESIDPHILV 204  
 356 ---LKLIVVHADQATVQCIAMNDVSGEASQAILV 388

RESULT 12  
 T14316  
 T14316 protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T14316  
 E:Y.Y.S.S.F.7 A. A. Busby, G. K. Lee, E.Y.H.P.  
 submitted to the EMBL data library, April 1998  
 A:Reference number: 217975  
 A:Accession: T14316  
 A:Status: preliminary; translated from cDNA/EMBL/TrEMBL  
 A:Molecule type: mRNA  
 A:Residues: 1-1344 (CDS)  
 A:Cross-references: EMBL:AF060570; NID:94206385; PDB:94206386; PDB:AA011620.1

Query Match 11.98; Score 154.5; DB 2; Length 1344;  
 Best Local Similarity: 24.74; Field No. 0.00075;  
 Matches 68; Conservative 38; Mismatches 76; Indels 47; Gaps 16.

9 LVLT---FLPREMTSSVLTNGKTEYILDTTPGASGLICAVQNTREBELMREBS 64  
 230 LVVLEPSPFLRPINOVLA-----DAEVNICEVQGP--QPNLHMRKDDG 274  
 65 RVLLKSN--KINS--SVGVSSISENLNISTFC---KLGKQSVSVYVNVNTPFL 117  
 275 --ELPACRFEIRSDHSMITIDVSSDEG--YTCVAENSVGRAN---SGSLSVHYTPQY 328  
 118 SGNPQVTEKSNVAKVYCNVKNANPQAMMYKNSLIDLEKSKHQLQUT-----TSSPOL 172  
 329 TPQNGVTAAPGANVSFGQETKGNPPAIFQKEGSGVLLPFG--SLPWRGLVSRGL 387  
 173 SLVYKTEK--NLYSTAKS---SLKTESLPHILVKKITV--GVYIEHII 217  
 388 NTFVKIGDPGYVYGVAVSAGSIARAL--LPIKASIDIGLP--PII 431

RESULT 13  
 T14316  
 protein tyrosine phosphatase (bc 3.1.3.48) D1AR precursor - fruit fly (Drosophila mel  
 N:Alternate names: leukocyte antigen-related protein  
 C:Species: Drosophila melanogaster  
 C:Date: 14 Dec 1990 #sequence\_revision 02 May 1994 #text\_change 22-Jun-1999  
 C:Accession: A36182  
 R:Stiehl, M.; Krieger, N.X.; Tsai, A.Y.M.; Saito, H.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 8698-8702, 1989  
 A:Title: A family of receptor linked protein tyrosine phosphatases in humans and Dros  
 A:Reference number: A36182; MIM:90046860  
 A:Accession: A36182





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:11:59 ; Search time 56.43 Seconds

(without alignments)  
446.135 Million cell updates/sec

Title: US-09-863-823 6

Perfect score: 1336

Sequence: 1 MAMKSSVIMOMGRFLLVIL.....RKTKMKCMKDKPHSETAL 262

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by 'chance' to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	176	13.2	853	1 IJBONC	neural cell adhesi
2	175	13.1	858	1 IJBONC	neural cell adhesi
3	173.5	13.0	2783	2 T34416	hypothetical prote
4	171.5	12.8	946	1 A47299	for-related recept
5	170	12.7	1896	2 T08851	Down syndrome cell
6	166	12.4	761	1 TJHUNG	neural cell adhesi
7	161	12.1	1011	2 T13669	neuroscullin - fr
8	160.5	12.0	1091	1 TJCHNL	neural cell adhesi
9	159	11.9	725	1 TJCHNL	neural cell adhesi
10	159	11.9	1115	1 TJCHNL	neural cell adhesi
11	156	11.7	1173	2 T25893	hypothetical prote
12	153.5	11.5	1344	2 T25893	hypothetical prote
13	153	11.5	2029	1 TCEFLK	protein-tyrosine p
14	152.5	11.4	5175	2 T20992	hypothetical prote
15	152.5	11.4	5198	2 T43290	hemocentin precurs
16	150	11.2	1912	2 A56178	protein-tyrosine p
17	149	11.2	1447	2 A54100	tumor suppressor p
18	148.5	11.1	338	2 JC4776	perlecan precursor
19	148.5	11.1	4391	2 A38096	perlecan precursor
20	146.5	11.0	773	2 T46283	hypothetical prote
21	146	10.9	7962	2 T38346	elastic titin - hu
22	144.5	10.8	338	2 JC5519	50K glycoprotein p
23	144.5	10.8	1348	2 S5156	vascular endotheli
24	144	10.8	458	2 WMSR1	billary glycoprote
25	144	10.8	521	2 JC1508	billary glycoprote
26	142.5	10.7	458	2 JC1509	billary glycoprote
27	142.5	10.7	521	2 S34338	billary glycoprote
28	142	10.6	871	1 T48696	protein-tyrosine k
29	142	10.6	881	1 T48697	protein-tyrosine k

30	142	10.6	1894	2 C54689	protein-tyrosine p
31	141	10.6	976	2 T29283	hypothetical prote
32	139	10.4	1450	2 A44027	155K myofibrillar
33	138.5	10.4	797	2 S38579	fibroblast growth
34	138	10.3	1880	2 T18531	tractin - medusa
35	137.5	10.3	1898	2 S46216	leukocyte antigen-
36	137	10.3	725	2 JC5105	neural cell adhesi
37	137	10.3	1092	1 JN0635	neural cell adhesi
38	136.5	10.2	458	2 S23969	cell-adhesion mole
39	136.5	10.2	458	2 S56177	C CAM2a protein is
40	136.5	10.2	519	2 A44783	ecto-ATPase precu
41	136.5	10.2	1897	1 TJCHNL	leukocyte antigen-
42	135	10.1	1273	2 T42405	sax-3 protein - Ca
43	135	10.1	1356	2 JC1462	protein-tyrosine k
44	134.5	10.1	1070	2 T44503	protein-tyrosine k
45	134	10.0	1375	2 T13822	frazzled gene prot

## ALIGNMENTS

RESULT 1

IJBONC

neural cell adhesion molecule short domain form precursor - bovine

N:Alternate names: NCAM-140

C:Species: Bos primigenius taurus (cattle)

C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 22-Jun-1999

C:Accession: A32976; A38778; B44290; S05402

R:Lipkin, V.M.; Khramtsov, N.V.; Andreeva, S.G.; Moshynakova, M.V.; Petukhova, G.V.; R

FEBS Lett. 254, 69-73, 1989

A:Title: Calmodulin-independent bovine brain adenylate cyclase Amino acid sequence a

A:Reference number: A32976; MIMD:89378239

A:Accession: A32976

A:Molecule type: mRNA

A:Residues: 1-853 <11P>

A:Cross-references: 5X; X16451; NTR-946; PDB:1CAA4470.1; PDB:961

A:Accession: A38778

A:Molecule type: protein

A:Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-5

A:Note: The authors identified this protein as calmodulin-independent adenylate cyclase

J. Biol. Chem. 261, 3396-3401, 1986

A:Title: Structural and immunological characterization of the amino-terminal domain o

A:Reference number: A44290; MIMD:86140120

A:Accession: B44290

A:Molecule type: protein

A:Residues: 20-36 <RDV>

A:Note: 23-Glu was also found

C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM

C:Comment: Various forms of NCAM are produced by alternative splicing.

C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; 1

C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;

F:119/Domain: signal sequence #status predicted <Sig>

F:20-85/Prodot: neural cell adhesion molecule short domain form #status experiment

F:30-719/Domain: extracellular #status predicted <EXT>

F:34-98/Domain: immunoglobulin homology <IM1>

F:132-191/Domain: immunoglobulin homology <IM2>

F:152-156/Region: heparin binding #status predicted

F:161-165/Region: heparin binding #status predicted

F:228-288/Domain: Immunoglobulin homology <IM3>

F:321-396/Region: NCAM binding #status predicted

F:428-490/Domain: immunoglobulin homology <IM4>

F:527-604/Domain: fibronectin type III repeat homology <FN1>

F:633-693/Domain: fibronectin type III repeat homology <FN2>

F:720-737/Domain: transmembrane #status predicted <TM>

F:738-853/Domain: intracellular #status predicted <INT>

F:412-96;139-159;245-280;328-394;435-488/Disulfide bonds: #status predicted

F:222;314;346;432;458;487/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 13.28, Score 176, DB 1, Length 833;

Best local Similarity 28.4%; Pred. No. 1e-05;

Matches: 50; Conservation: 31; Mismatches: 88; Indels: 32; Gaps: 8.

QY 49 QASLIC-----AVNHTREELMYREGRDLKSGNKINS-----SSVCYS 90  
 ID 325 QVTLTGASGQIPSPSLMRISTENISLENASWLRPNQVTLQDHMYVNSHAAVSSLLIK 184  
 QY 91 SISENDINGISPTG-----RLGRDQSVSVVLTNTPPLISQNDFOVESSNKLIVNWK 146  
 ID 385 SIQYTDAG-EYICIASNTIGD-----SOSMTLEVOYAPKLDGPVAVYTMENQVNTCEVF 440  
 QY 147 ANPQAMMMYNNSSLDLQEK-SFRLQGLSPSPULIKKPKPLNLTYSGLASSLKLES 205  
 ID 439 AVPSATISWFRGQGLPSSNYSNLIKTYNTPSASTLEVPSENFPGVNTAVRIGQES 498  
 QY 206 LDFELIVKDKYGVF-----LEFLIACVVF 242  
 ID 499 LEFLIVQAD-TPSSPIDRVEPYSTACQVF 528

RESULT 2

Neutral cell adhesion molecule short domain form precursor - rat

N:Altorfate names: NCAM-140  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1991 \*sequence\_revision 10-Sep-1991 \*text\_change 22-Jun-1999  
 C:Accession: S00846; GI:7795; EMBL:158136  
 R:Small, S.J.; Shull, G.F.; Santoni, M.J.; Akesson, K.  
 J:Cell Biol. 105, 2345-2349, 1987  
 A:Title: Identification of a cDNA clone that contains the complete coding sequence for a  
 A:Reference number: S00846; MUID:8605265  
 A:Accession: S00846  
 A:Molecule type: mRNA  
 A:Residues: 1-858 <MA>  
 A:Cross references: EMBL X65664  
 R:Small, S.J.; Akesson, K.  
 J:Cell Biol. 111, 2087-2090, 1990  
 A:Title: Expression of the unique NCAM VASE exon is independently regulated in distinct  
 A:Reference number: A37795; MUID:91035620  
 A:Accession: M37796  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 340-481 <SM2>  
 R:Small, S.J.; Haines, S.L.; Akesson, K.A.  
 Neuron 1, 1007-1017, 1988  
 A:Title: Polypeptide variation in an N-NCAM extracellular immunoglobulin like fold is dev  
 A:Reference number: 158136; MUID:90166485  
 A:Accession: 158136  
 A:Status: preliminary; translated from cDNA/EST/EST  
 A:Molecule type: DNA  
 A:Residues: 455-464 <RES>  
 A:Cross references: GI:612611, M10-920543, F108.AA1:679.1, F10-920544  
 A:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol  
 A:Comment: Various forms of NCAM are produced by alternative splicing.

Accession: NCAM

C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu  
 C:Keywords: alternative splicing; brain; cell adhesion; duplicated; heparin binding; su  
 F:1-19/Domain: signal sequence \*status predicted <SIG>  
 F:20-858/region: neural cell adhesion molecule, short domain form \*status predicted <MA  
 F:20-858/region: neural cell adhesion molecule, short domain form \*status predicted <MA  
 F:44-48/Domain: extracellular \*status predicted <EXT>  
 F:44-48/Domain: immunoglobulin homology <IMM1>  
 F:142-191/Domain: immunoglobulin homology <IMM2>  
 F:152-159/region: heparin binding \*status predicted  
 F:161-165/region: heparin binding \*status predicted  
 F:228-290/Domain: immunoglobulin homology <IMM3>  
 F:263-272/Region: NCAM binding \*status predicted  
 F:323-398/Domain: immunoglobulin homology <IMM4>  
 F:430-492/Domain: immunoglobulin homology <IMM5>  
 F:529-605/Domain: fibronectin type III repeat homology <FN3A>  
 F:635-695/Domain: fibronectin type III repeat homology <FN3B>  
 F:722-739/Domain: transmembrane \*status predicted <TM>  
 F:740-858/Domain: intracellular \*status predicted <INT>  
 F:41-96,139-189,245-288,330-396,437-490/Cisaltide bonds: \*status predicted

1582,316,348,431,460,489, binding site, carbohydrate (Asn) (conserved) \*status predict

Query Match 13.18; Score 175; DB 1; Length 858;  
 Best local similarity: 28.48; Pred. No. 1 20-05;  
 Matches: 60; Conservation: 31; Mismatches: 88; Indels: 32; Gaps: 8;

QY 49 QASLIC-----AVNHTREELMYREGRDLKSGNKINS-----SSVCYS 90  
 ID 325 QVTLTGASGQIPSPSLMRISTENISLENASWLRPNQVTLQDHMYVNSHAAVSSLLIK 184  
 QY 91 SISENDINGISPTG-----RLGRDQSVSVVLTNTPPLISQNDFOVESSNKLIVNWK 146  
 ID 385 SIQYTDAG-EYICIASNTIGD-----SOSMTLEVOYAPKLDGPVAVYTMENQVNTCEVF 440  
 QY 147 ANPQAMMMYNNSSLDLQEK-SFRLQGLSPSPULIKKPKPLNLTYSGLASSLKLES 205  
 ID 439 AVPSATISWFRGQGLPSSNYSNLIKTYNTPSASTLEVPSENFPGVNTAVRIGQES 498  
 QY 206 LDFELIVKDKYGVF-----LEFLIACVVF 242  
 ID 499 LEFLIVQAD-TPSSPIDRVEPYSTACQVF 528

RESULT 3

Hypothetical protein, F12F3.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 \*sequence\_revision 29-Oct-1999 \*text\_change 29-Oct-1999  
 C:Accession: T34416  
 R:Fulton, B.; Wohldmann, P.  
 submitted to the EMBL Data Library, July 1998  
 A:Description: The sequence of C. elegans cosmid F12F3.  
 A:Reference number: 231521  
 A:Accession: T34416  
 A:Status: preliminary; translated from cDNA/EST/EST  
 A:Molecule type: DNA  
 A:Residues: 1-2783 <FULL>  
 A:Cross references: EMBL D00662; F108.AA:25806.1; GSI08-000023; FSP:F12F3.2  
 A:Experimental source: strain Bristol N2; clone F12F3  
 C:Genetics:  
 A:Gene: CESP-F12F3.2  
 A:Map position: 5  
 A:Introns: 45, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000

for-related receptor RPK - Pacific electric ray  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.12)  
 C:Species: Torpedo californica (Pacific electric ray)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-May-2000  
 C:Accession: A47299  
 R:Jennings, C.G.; Dyer, S.M.; Burden, S.J.  
 Proc. Natl. Acad. Sci. U.S.A. 90: 2895-2899, 1993  
 A:Title: Muscle-specific trk-related receptor with a kringle domain defines a distinct  
 A:Reference number: A47299; MIM:93219391  
 A:Accession: A47299  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-946 <1EN>  
 A:Cross-references: GB:U11311; NID:q290857; PIDN:AAA9285.1; FID:q290858  
 A:Experimental source: electric organ  
 A:Note: sequence extracted from NCBI backbone (NCBIN:128724, NCBI:128726)  
 C:Superfamily: Torpedo for-related receptor; immunoglobulin homology; kringle homology;  
 C:Keywords: ATP; glycoprotein; kringle; phosphotransferase; transmembrane protein; tyro  
 F:44-103/Domain: immunoglobulin homology <IM1>  
 F:137-195/Domain: immunoglobulin homology <IM2>  
 F:229-287/Domain: immunoglobulin homology <IM3>  
 F:464-542/Domain: kringle homology <KR3>  
 F:572-588/Domain: transmembrane #status predicted <1MN>  
 F:650-940/Domain: protein kinase homology <KIN>  
 F:658-666/Region: protein kinase ATP-binding motif  
 F:225,340,477,544/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 12.8%; Score 171.5; DB 1; Length 946;  
 Best Local Similarity 25.1%; Pred. No. 2.5e-05;  
 Matches 63; Conservative 38; Mismatches 95; Indels 55; Gaps 10;

QY 15 LLLVIL-----FLPR--EMTSSVLTVNGKTEVITDTPGASALICAVQNTREE 63  
 Db 10 LLMFLVLTGGSADGILPKAPVLTISPLEVAVLVE-----EASPMCAVDSYR-AA 59  
 QY 64 ELTVREGRV---DLKSNKINSSVCSVSSISNDNGI-SFTCRGRQSVSVYLVNV 119  
 Db 60 EITWRNNIPRPDTRYSKENGQILTILSVEDTDNGVCCCRANNMGSSASQCALAV 119  
 QY 120 TFPF--LLSGNPFQTEEGSNVNLVKNPQACMMKYNSSLDELSPHIQQTSES 177  
 Db 120 KMKPKIIRPPIDVRL-IGSKVVLPCSTMCNPKPAISWFKDELALNKDPRISV--LBS 175  
 QY 178 FOLSTVEXKPDNGTSCIAKSSIKTE-----SLDFLIYK 213  
 Db 176 GNLKIRNVQLEDACKYKCLARNSLGFYSRSLAEVQVSARIVKAPTSONVSTGSEVILQ 235  
 QY 214 DRTGVPIEP1 224  
 Db 216 CKATGFPIPT1 246

RESULT 5  
 T08851  
 Down syndrome cell adhesion protein 1 - human (fragment)  
 N:Alternate names: Down syndrome cell adhesion molecule  
 C:Species: Homo sapiens (man)  
 C:Date: 11-Jun-1995 #sequence\_revision 11 Jun 1999 #text\_change 11 Jun 1999  
 C:Accession: T08851  
 R:Yanaka, K.; Huo, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.F.; Korenberg  
 submitted to the EMBL Data Library, September 1997  
 A:Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down sy  
 A:Reference number: 216495  
 A:Accession: T08851  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1896 <YAM>  
 A:Cross-references: EMBL:AF023449; NID:q3169765; FID:q3169766  
 A:Experimental source: brain; developmental stage: 14 weeks; fetal  
 C:Genetics:  
 A:Gene: DSCAM  
 A:Map position: 21q22

A:Note: derived from alternatively spliced mRNA  
 C:Function:  
 A:Description: involved in nervous system development  
 C:Keywords: alternative splicing

Query Match 12.7%; Score 170; DB 2; Length 1896;  
 Best Local Similarity 29.8%; Pred. No. 7.6e-05;  
 Matches 57; Conservative 31; Mismatches 69; Indels 34; Gaps 10;

QY 23 PREMTSSVLTVNGKTEVITDTPGASALICAVQNTREEELWYREGVLDLKGK- 81  
 Db 306 PRKXSSV-----GSOVLSGCVTG-TEDELSTWR-NGEL-LMPGKNV 346  
 QY 82 ---INSSVCSVSSISNDNGISFTCRGRQSVS--SVVLNVTFPLLSGNDPQVE 134  
 Db 347 RITGINENILMDHMYKSDG-AVQCFVRKDKLSADVYQVLEDDTPKLTISAFSEKVS 405  
 QY 135 EGSNVKLVCKVKNPQACMMKYNSSLDELKSRHQLQ--TSES--FOLSTVEXKPD 189  
 Db 406 PAEVSILCKVKGTFLEFTITWTLGLCFI-LKGSKRISQMTISGWNVSYLMTSSVYRC 464  
 QY 190 NQTSYCLAKSS 400  
 Db 465 GGVRCTANNNS 475

## RESULT 6

IJHUNG  
 neural cell adhesion molecule 1 GPI anchored splice form precursor, muscle specific -

N:Alternate names: CD56; NCAM-120

C:Species: Homo sapiens (man)  
 C:Date: 31-Mar-1994 #sequence\_revision 31-Mar-1994 #text\_change 28-Jan-2000

C:Accession: S07784; A26883  
 R:Barton, C.H.; Dickson, G.; Gower, H.J.; Kowett, L.H.; Pull, W.; Elsom, V.; Moore, S

development 104, 165-173, 1988  
 A:Title: Complete sequence and in vitro expression of a tissue-specific phosphatidyl

A:Reference number: S07784; MIM:49495258

A:Accession: S07784

A:Molecule type: mRNA

A:Residues: 1-761 <BNV>  
 A:Cross-references: EMBL:X16841; NID:q35005; PIDN:CAA34739.1; PID:q35006

R:Dickson, G.; Gower, H.J.; Barton, C.H.; Prentice, H.M.; Elsom, V.L.; Moore, S.E.; C

ell 50, 1119-1130, 1987

A:Title: Human muscle neural cell adhesion molecule (N CAM). Identification of a musc

A:Reference number: A90895; MIM:8730755

A:Accession: A26883

A:Molecule type: mRNA

A:Residues: 491-761 <DLQ>

A:Cross-references: GB:M17409; NID:q189097; PIDN:AAA59912.1; PID:q386979

C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM

C:Comment: Various forms of NCAM are produced by alternative splicing.

C:Genetics:

A:Gene: GDB:NCAM1; NCAM; CD56

A:Cross-references: GDB:119448; OMIM:116930

A:Map position: 11q22.2-11q22.3

C:Superfamily: neural cell adhesion molecule, fibronectin type III repeat homology; i

C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membra

F:113/Domain: signal sequence #status predicted <SIG>  
 F:20-761/Product: neural cell adhesion molecule phosphatidylinositol-linked form, mus  
 F:134-98/Domain: immunoglobulin homology <IMM1>  
 F:113-191/Domain: immunoglobulin homology <IMM2>  
 F:154-156/Region: heparin binding #status predicted  
 F:161-165/Region: heparin binding #status predicted  
 F:228-289/Domain: immunoglobulin homology <IMM3>  
 F:263-272/Region: NCAM binding #status predicted  
 F:322-387/Domain: immunoglobulin homology <IMM4>  
 F:419-481/Domain: immunoglobulin homology <IMM5>  
 F:493-587/Domain: fibronectin type III repeat homology #status atypical <FN3A>  
 F:633-720/Domain: fibronectin type III repeat homology #status atypical <FN3B>  
 F:41-96,139-189,329-385,426-479/Disulfide bonds: #status predicted  
 F:222,315,347,423,443,478/Binding site: carbohydrate (Asn) (covalent) #status predict

Only Match 161, 12-18, Score 161, DB 2, Length 1011, Host Local Similarity 24-48, Freq. No. 6-00018, Matches 46, Unsuccessful 34, Mismatches 81, Indels 26, Gaps 6.

```
0Y 41 LTVNGKNTNYILDTTPGQVASTLCAVQNTFREELMWR-----EGRVDL-----76
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 265 IDINCKIKH-----TVQSGSVLLDIDHAKPAVNLFWNTILSSGENLELVKRSKSL 320
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 77 KSNKINSSSVYSSISSENGISPT-----PCHQSVSVVNLNTPPLSNP 129
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 421 KSNKTHQSELFNATRENDPVECEANIVLOINRKLSSATLEVLPPVKKVSP 380
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 140 PVLVEGGSNVLYTNVKNAPCA--QMWYANSSILDEKSHQVQVTSFQSLVYR 186
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 381 SATANTSEIVLLNCEYANPASTGVPEYRNDLVNVNFTTHYKQNSFNVAIVKSTF 440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 187 KPLNGTYSQ 195
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 441 KEDIDNYSQ 449

RESULT R
LIGNIN.
neural cell adhesion molecule long domain form precursor - chicken
NCAltorate names: NCAM-180
NCContalust: neural cell adhesion molecule, short domain form (NCAM-140)
CSpecies: Gallus gallus (chicken)
CDate: 41 Mar 1994 *sequence revision, 41 Mar 1994 *ext change 22-10-1994
CAccession: A43613; E43613; A25435; B25435; A46550; S36950; A44369; A60852; S29668
RCUnlqham, B.A.; Hempeley, J.D.; Murray, B.A.; Prediger, E.A.; Brackenbury, R.; Edelfelt,
Science 236, 793-806, 1987
```

A>Title: Neuronal adhesion molecule; structure, immunoglobulin-like domains, cell  
A:Reference number: A43613; MUID:87296190

A:Accession: A43613  
A:Molecule type: mRNA  
A:Residues: 1-175 <CD>  
A:Cross-references: GR:M15860  
A:Accession: A43613

A:Molecule type: protein  
A:Residues: 20-44,126-127,202-201,320-342,399-415,440-459,822-828,~C938<  
A:Note: Asp 222 probably binds carboxylate; Asp 226 probably does not.  
Proc. Natl. Acad. Sci. U.S.A. 83, 3037-3041, 1986  
A:Title: Sequence of a cDNA clone encoding the polysialic acid-rich and cytoplasmic d  
A:Reference number: A25435; MUID:86206089

A:Accession: A25435  
A:Molecule type: mRNA  
A:Residues: 128-1091 <HEM>

A:Cross-references: GR:M13210  
A:Accession: B25435

A:Molecule type: protein  
A:Residues: 128-140,127-240,428-439,554-563,744-761,764-781,1080-1084,~H22<  
J. Murray, B. A.; Owens, G.C.; Preidiger, E. A.; Crossin, K.L.; Cunningham, B. A.;  
Edelman, J. Cell Biol. 103, 1431-1439, 1986

A:Title: Cell surface modulation of the neural cell adhesion molecule resulting from  
A:Reference number: A46550; MUID:87033934

A:Accession: A46550  
A:Molecule type: DNA  
A:Residues: 810-1070 <MOR>

A:Cross-references: GR:X04479  
R. Sasser, M.; Covault, J.  
submitted to the EMBL Data Library, February 1993

A:Reference number: S36950  
A:Accession: S36950

A:Molecule type: DNA  
A:Residues: 1-17 <SAS>

A:Cross-references: EMBL:X70342; NID:q417631; PID:CAA49807.1; PID:q417632  
P. C. O'Neil, G. H. R. Forrest, D. Brackenbury, K.  
Genomics 14, 875-882, 1992

A:Title: Conserved regulatory elements in the promoter region of the NCAM gene.  
A:Reference number: A44368; MUID:93122797

A:Accession: A44369  
A:Molecule type: DNA  
A:Residues: 1-17 <SCB>

A:Cross-references: EMBL:212129; NID:q34543; PIRN:VMA74113.1; PIR:q34554  
A. Experimental source: White leghorn

R. Cole, G.J.; Loewy, A.; Gross, N.V.; Akesson, K.; Glaser, L.  
J. Cell Biol. 103, 1739-1744, 1986

A:Title: Topographic localization of the heparin-binding domain of the neural cell ac  
A:Reference number: A60852; MUID:87057627

A:Accession: A60852  
A:Molecule type: protein  
A:Residues: 20-29 <COE>

F. Rao, Y.; Wu, X.F.; Garlepy, J.; Rutishauser, U.; Stu, C.H.  
F. Cell Biol. 118, 937-949, 1992

A:Title: Identification of a peptide sequence involved in hemophilic binding in the r  
A:Reference number: A43280; MUID:92763934

A:Contents: annotation; homophilic binding region  
A:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM  
C:Comment: Various forms of NCAM are produced by alternative splicing.  
C:Keywords: neural cell adhesion molecule; fibronectin type III repeat homology; h  
C:Keywords: alternative splicing; brain; cell adhesion; heparin binding; ~IM45<  
F.1-19/Domains: signal sequence \*status predicted <SI16>  
F.20-1091/Product: neural cell adhesion molecule, long domain form \*status experimental  
F.20-809,1071-1091/Product: neural cell adhesion molecule, short domain form \*status  
F.20-111/Domains: extracellular \*status predicted -EXT<  
F.32-98/Domains: immunoglobulin homology <IMH1>  
F.132-191/Domains: immunoglobulin homology <IMH2>  
F.152-156/Region: heparin binding \*status predicted  
F.161-165/Region: heparin binding \*status predicted  
F.~R,184/-main: immunoglobulin homology -IMH3<  
F.262-271/Region: NCAM binding \*status experimental  
F.322-387/Domains: immunoglobulin homology <IMH4>  
F.415-481/Domains: immunoglobulin homology <IMH5>



F:518-595/Domain: fibronectin type III repeat homology <FN3A>  
 F:624-685/Domain: fibronectin type III repeat homology <FN3B>  
 F:712-724/Domain: transmembrane \*status predicted <TM>  
 F:770-1091/Domain: intracellular \*status predicted <INT>  
 F:41-96,139,235,287,329-385,426,479/Disulfide bonds: \*status predicted  
 F:222/Binding site: carbohydrate (Asn) (covalent) \*status experimental  
 F:315,347,423,445,478/Binding site: carbohydrate (Asn) (covalent) \*status predicted

Query Match 12.0%; Score 160.5; DB 1; Length 1091;  
 Best Local Similarity 26.9%; Pred. No. 0.00021;  
 Matches 60; Conservative 37; Mismatches 87; Indels 39; Gaps 10;

```

OY 33 VNGTENTIIDTTCGQASLIC-----AVQNHTEBELLMYREGGRDLSK 78
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 313 VENTAMELED-----DTTLCEASGDPPIPTWKTSTRNISNEETL----DGRIVRS 363
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
OY 79 GNKINSSVCYSISENDNGISFTC---RLGRQSVSVVLTNPPLLSGNDFOVE 134
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 364 HARV--SSLTLEKEIOTYDAG--EYVCTASNTIGD---SQAMYLEVOYAPKLGSPAVYTM 417
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
OY 135 EGSNVKLVCKNKPQAKQMMYKNSLLDLEK--SRHQIQGTSESFQSLSTKVEKPNQGY 193
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 418 EGNQVNITCEVFAYPSAVISWFRDQLPSSNYSNLIKITYNPASYLEVTPDSEDFGMY 477
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
OY 194 SCIAKSSIKTESLDFHLIVCKITGVF----IEPIIACVYIF 232
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 478 NCTAVNRIGQPSSEFLLVQAD--TPSSPSIDRVEYSSTARVEF 519
   ||| | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 9

neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse

NCAM-120  
 C:Species: Mus musculus (house mouse)  
 C:Date: 31-Mar-1993 #sequence, revision 31-Mar-1993 #text, change 31-Dec-2000  
 C:Accession: A29673; S00382; A44290  
 R:Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fontec  
 EMBL J. 6, 907-914, 1987  
 A>Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000  
 A:Reference number: A29673; MUID:87246524  
 A:Accession: A29673  
 A:Molecule type: mRNA  
 A:Residues: 1-725 <BAR>  
 A:Cross-references: EMBL:Y00051; NID:953342; PIDD:CAA68263.1; PTD:953343  
 R:Barthels, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.  
 EMBL J. 7, 625-632, 1988  
 A>Title: Differential splicing and alternative polyadenylation generates distinct NCAM  
 A:Reference number: S00382; MUID:88283628  
 A:Accession: S00382  
 A:Molecule type: DNA  
 A:Residues: 642-656, 'T', 658-735 <RA>  
 A:Cross-references: EMBL:X07195  
 R:Rougon, G.; Marshak, D.R.  
 J. Biol. Chem. 261, 3396-3401, 1986  
 A>Title: Structural and immunological characterization of the amino-terminal domain of m  
 A:Reference number: A44290; MUID:86140120  
 A:Accession: A44290  
 A:Molecule type: protein  
 A:Residues: 20-36 <ROU>  
 C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol  
 C:Comment: Several forms of NCAM are produced by alternative splicing. See also PIR, IJMS  
 C:Genetics:  
 A:Gene: NCAM  
 A:Map position: 9  
 A:Introns: 701/1  
 C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu  
 C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane  
 F:1-19/Domain: signal sequence \*status predicted <SIG>  
 F:34-98/Domain: immunoglobulin homology <IMM1>  
 F:132-191/Domain: immunoglobulin homology <IMM2>  
 F:152-156/Region: heparin binding \*status predicted  
 F:161-165/Region: heparin binding \*status predicted

F:418-518/Domain: immunoglobulin homology <IMM3>  
 F:263-272/Region: NCAM binding \*status predicted  
 F:323-388/Domain: immunoglobulin homology <IMM4>  
 F:426-482/Domain: immunoglobulin homology <IMM5>  
 F:519-596/Domain: fibronectin type III repeat homology <FN3A>  
 F:625-685/Domain: fibronectin type III repeat homology <FN3B>  
 F:41-96,139,189,235,288,330-386,427,480/Disulfide bonds: \*status predicted  
 F:222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) \*status predict

Query Match 11.9%; Score 159; DB 1; Length 725;  
 Best Local Similarity 27.5%; Pred. No. 0.00017;  
 Matches 57; Conservative 35; Mismatches 81; Indels 34; Gaps 9;

```

OY 49 QASLIC-----AVQNHTEBELLMYREGGRDLSKGNKINSSVCYSISE 94
   ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 325 QVTLTCEASGDPPIPTWKTSTRNISSEEDL----DGRVVRSHARV--SSLTLEKIUY 378
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
OY 95 NDNGISFTC---RLGRQSVSVVLTNPPLLSGNDFOVEEGSNVCKLVCKNKPQ 150
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 379 KDAG--EYVCTASNTIGD---SQAMYLEVOYAPKLGSPAVYTMEDQVNTICEVFAPYS 434
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
OY 151 AQMMYKNSLLDLEK--SRHQIQGTSESFQSLSTKVEKPNQGYSCIAKSSIKLSLDFH 209
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 435 ATISWFRDQLPSSNYSNLIKITYNPASYLEVTPDSEDFGMYNCTAVNRIGQPSSEF 494
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
OY 210 LTVKDKTIVGVF----IEPIIACVYIF 232
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 495 LVQAD--TPSSPSIDRVEYSSTARVEF 520
   ||| | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 10

neural cell adhesion molecule 1 precursor, long domain splice form mouse

NCAM-180  
 C:Species: Mus musculus (house mouse)  
 C:Date: 31-Mar-1993 #sequence, revision 31-Mar-1993 #text, change 31-Dec-2000  
 C:Accession: A29673; S00844; S00382; A44290; S00383  
 R:Barthels, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fon  
 EMBL J. 6, 907-914, 1987  
 A>Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,  
 A:Reference number: A29673; MUID:87246524  
 A:Accession: A29673  
 A:Molecule type: mRNA  
 A:Residues: 1-548, 'T', 550-571, 'T', 573-574, 'D', 576-588, 'MQPS', 593, 'S', 595-599, 'P', 601,  
 A:Cross-references: EMBL:Y00051; NID:953342; PIDD:CAA68263.1; PTD:953343  
 R:Santoni, M.J.; Barthels, D.; Barthels, J.A.; Hirsch, M.R.; Steinmetz, M.; Goridis, C.  
 Nucleic Acids Res. 15, 8621-8641, 1987  
 A>Title: Analysis of cDNA clones that code for the transmembrane forms of the mouse n  
 A:Reference number: S00844; MUID:88067687  
 A:Accession: S00844  
 A:Molecule type: DNA  
 A:Residues: 529-809, 1077-1115 <SAN>  
 A:Cross-references: EMBL:X06328; NID:953322; PIDD:CAA29641.1; PTD:9817984  
 R:Barthels, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.  
 EMBL J. 7, 625-632, 1988  
 A>Title: Differential splicing and alternative polyadenylation generates distinct NCA  
 A:Reference number: S00382; MUID:88283628  
 A:Accession: S00382  
 A:Molecule type: DNA  
 A:Residues: 642-1115 <BAR>  
 A:Cross-references: EMBL:X07195  
 R:Barthels, D.; Vopper, G.; Wille, W.  
 Nucleic Acids Res. 16, 4217-4225, 1988  
 A>Title: NCAM-180, the large isoform of the neural cell adhesion molecule of the mous  
 A:Reference number: A44281; MUID:88247737  
 A:Accession: A44281  
 A:Molecule type: mRNA  
 A:Residues: 804-1081 <BA3>  
 A:Cross-references: EMBL:X07244; NID:953321; PIDD:CAA68263.1; PTD:9329720  
 R:Rougon, G.; Marshak, D.R.  
 J. Biol. Chem. 261, 3396-3401, 1986

A: Accession: A44290  
 A: Molecule type: protein  
 A: Residues: 207-46 (460)  
 C: Comment: NCAM mediated cell adhesion via homophilic binding with another NCAM molecule.  
 C: Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:J03603.  
 A: Gene: NCAM  
 A: Map position: 9  
 A: Ratios: 64/3: 70/1: 770/2: 809/2: 1076/2  
 E: Superfamily: neural cell adhesion molecule, fibronectin type III repeat homology, immunoglobulin  
 E: Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; signal  
 E: 1-19/Domain: signal sequence; \*status predicted<EXT>  
 E: 20-115/Product: neural cell adhesion molecule, long domain splice form \*status expert  
 E: 207-809/107-115/Product: neural cell adhesion molecule, short domain splice form \*status  
 E: 207-711/Domain: extracellular \*status predicted<EXT>  
 E: 320-98/Domain: immunoglobulin homology<IMM1>  
 E: 132-191/Domain: immunoglobulin homology<IMM2>  
 E: 152-156/Region: heparin binding \*status predicted  
 E: 161-165/Region: heparin binding \*status predicted  
 E: 228-290/Domain: immunoglobulin homology<IMM3>  
 E: 262-272/Region: NCAM binding \*status predicted  
 E: 323-388/Domain: immunoglobulin homology<IMM4>  
 E: 420-482/Domain: immunoglobulin homology<IMM5>  
 E: 519-596/Domain: fibronectin type III repeat homology<FN3A>  
 E: 625-685/Domain: fibronectin type III repeat homology<FN3B>  
 E: 712-729/Domain: transmembrane \*status predicted<TM>  
 E: 730-1115/Domain: intracellular \*status predicted<IN>  
 E: 741-96,139-189,235-388,430-386,427-480/Disulfide bonds: \*status predicted  
 E: 222,316,348,424,450,479/Binding site: cytochrome (AS0) (conserved) \*status predicted

Query Match 11.9% Score 159; DB 1; length 1115;  
 best local similarity 27.5% Freq. No. 3,60028,  
 Matches 57; conservative 45; Mismatches 81; Indels 34; Gaps 9;

QY 49 QASL... AV...  
 DB 425 QVTLTGRSGSDPPISTWTSTPNISFEQD...DDHMYRSHARV...SSTLTKSIQY 378  
 QY 95 NINGST...KAKIKSVSVSVLWVTFEPHLSGNDLPVLEGSVWKLKCNVANK... 150  
 DB 379 RIAG-EYMGTAASNTIGD...SOSTDEFGYAPKIQGVAVYTWGNGVNITGEVAFPS 434  
 QY 151 AVMHAKKNSLILDLAK...SKVGLVQVLSHPVQLSLNVAEPQWCTISGTLAKSSIKTSLDFH 209  
 DB 435 ATTSWPKMGLDLPSSVSNRIKYNFASYLEVTDPDSNDGNGNCTAVNRIGQSELEFI 494  
 QY 210 LTVKDFTEGV...LEPILACVTF 232  
 DB 495 LVQAD-TPSSPTLQKVEPYSTAGVQF 520

RESULT 11  
 T25893  
 Description: hypothetical protein T1984.7, Caenorhabditis elegans  
 C: Species: Caenorhabditis elegans  
 C: Date: 15-Oct-1999 \*sequence revision 15 Oct 1999 \*text change 15-Oct-1999  
 C: Accession: T25893  
 E: Ratios: 5:  
 submitted to the EMBL Data Library, November 1996  
 A: Description: The sequence of C. elegans cosmid T1984.  
 A: Reference number: Z20186  
 A: Accession: T25893  
 A: Status: preliminary; translated from cDNA/EMBL/DDBJ  
 A: Molecule type: DNA  
 A: Residues: 1-1173 (GAT)  
 A: Sequences referred to: EMBL:080418, F0108 AAA37635.1; CSEPB:GN00619; CSEST:T1984.7  
 A: Experimental source: strain Bristol N2; clone T1984  
 A: Accession:  
 A: Gene: CSEST:T1984.7  
 A: Map position: 1

A:Accession: A36182  
 Query Match 11.7% Score 156; DB 2; Length 1173;  
 Best Local Similarity 26.4%; Pred. No. 0.00051;  
 Matches 57; Conservative 38; Mismatches 85; Indels 36; Gaps 10;

0Y 15 LILVILFLPRETSSVLTVNCKENTYLIOTDP-----GSGASLFCVAVNTPREELM 67  
 1B 191 LILAKTKLQKQDFSETV-----SNELVFTTRHLDGVNQGDFELKCLVASILR-POYRW 244  
 0Y 68 YEEGR VDKRSGKRINS3SVVSSISSENDGISTEEL GRPQSVASVAVINATPEPL 124  
 1B 245 LKESWLLVLELVYKIEFVSSILVSKASLLEGL-YLEFASNNQSLDRAVSEVAPPP 303  
 0Y 125 LSGN+LTVF--TSVKVW--NRKANG+QVMWYNN-----SSILQFA-SPLGLTSLF 176  
 1B 304 LTRPTTKVANETADVELECGTAARPERARWAKNGEATIGSEYFVLEPNR----- 355  
 0Y 177 SFQSLITVEKPDNGTYSGLANSKTESLPEHLIT 212  
 1B 356 ---LRLGIVADQDAIYCICIAENGVSQKASPOLIV 388  
 DB

RESULT 12  
 T14316  
 rig-1 protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 26 Sep 1993  
 C:Accession: T14316  
 F:From: G.S.F. Co., L.A., Calif., G.K.J. Lee, E.Y.H.P.  
 Submitted to the EMBL Data Library, April 1998  
 A:Reference number: Z17975  
 A:Accession: T14316  
 A:Status: Preliminary; Isolated from CB/TML/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1344 <YDA>  
 A:Accession: T14316  
 A:From: G.S.F. Co., L.A., Calif., G.K.J. Lee, E.Y.H.P.  
 Submitted to the EMBL Data Library, April 1998

Query Match 11.5% Score 153.5; DB 2; Length 1444;  
 Best Local Similarity 29.7%; Pred. No. 0.00095;  
 Matches 68; Conservative 38; Mismatches 76; Indels 47; Gaps 16;

0Y 17 LVLIL----FLPKREKISSVLTVNCKENTLIOTDPSGASLFCVAVNTHLETFITWYPPFG 72  
 1B 210 LVLILPSPSTLRPPINQVVA-----DAPNPFCEVGDP-QNLMRKKDG 274  
 0Y 73 RYDEESGN-KINSS-SVCVSSISENDGISTEEL RIGQDSVSVAVINATPEPL 125  
 1B 275 --LTPAGKTHLSHSLMLDQVSSSLDQ-ITICVAFNSVGAFA---SGSLSVHYPPQPV 328  
 0Y 125 LSGN+LTVF--TSVKVW--NRKANG+QVMWYNN-----SSILQFA-SPLGLTSLF 176  
 1B 329 TKPQMDTAAFGANVSFOGCTKGNPPALFMWKEGSGVILFSSQ-SLOPMGRILVSPKQL 387  
 0Y 181 SLLEFTEKNGISTEELSSLSLSLHLFKNLTV-GVLEPPI 225  
 1B 388 NITEVKIGDGGTYVQAVSVASGLAKAL---LETKASLIDSLP--PII 431  
 DB

RESULT 13  
 TDFPK  
 protein-tyrosine phosphatase (DC 3.1.3.48) BLAR precursor fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 14 Dec 1990  
 C:Accession: A36182  
 F:From: M.J. Kreiger, N.X. Tsai, A.Y.M. Saito, H. Proc. Natl. Acad. Sci. U.S.A. 86, 8998-9002, 1989  
 A:Title: A family of receptor-linked protein tyrosine phosphatases in humans and Dros.  
 A:Reference number: A36182; MUID:90040860  
 A:Accession: A36182





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:10:53 ; Search time 40.16 Seconds

(without alignments)  
244.890 Million cell updates/sec

Title: us-09-863-823-2

Perfect score: 1295

Sequence: 1 MQMGRFLLVILFLPRDNTS.....RKIKMLCKKPKDPSRTAL 254

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

105224

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176	13.6	853	1 NCAL_BOVIN	P1836 Bos taurus
2	175	13.5	858	1 NCAL_RAT	P13596 ratius norv
3	170	13.1	2012	1 DSCA_HUMAN	O60469 homo sapien
4	169.5	13.1	359	1 LACH_DROME	O24372 drosophila
5	166	12.8	761	1 NCAL_HUMAN	P13597 homo sapien
6	166	12.8	848	1 NCAL_HUMAN	P13597 homo sapien
7	160.5	12.4	1091	1 NCAL_CHICK	P13596 galus gall
8	159	12.3	725	1 NCAL_MOUSE	P13594 mus musculu
9	159	12.3	1115	1 NCAL_MOUSE	P13594 mus musculu
10	153	11.8	2029	1 LAR_DROME	P16621 drosophila
11	149	11.5	1447	1 DCC_HUMAN	P43146 homo sapien
12	149	11.5	1912	1 PPRD_HUMAN	P23466 homo sapien
13	148.5	11.5	338	1 LAMP_HUMAN	O13448 homo sapien
14	147.5	11.4	4393	1 PGBM_HUMAN	P98160 homo sapien
15	147	11.4	1447	1 DCC_MOUSE	P70211 mus musculu
16	145	11.2	837	1 NCAL_HUMAN	O15394 homo sapien
17	144.5	11.2	338	1 LAMP_CHICK	O98919 galus gall
18	144.5	11.2	338	1 LAMP_RAT	O62813 ratius norv
19	144.5	11.2	1348	1 VGR2_COLIA	P52585 colutrix cu
20	144	11.1	521	1 CEAL_MOUSE	P31809 mus musculu
21	142.5	11.1	149	1 LACH_HUMAN	O26474 s-hist-1-1
22	142	11.0	837	1 NCAL_MOUSE	O35136 mus musculu
23	142	11.0	1356	1 VGR2_HUMAN	P59696 homo sapien
24	141.5	10.9	1493	1 NCAL_MOUSE	O02177 galus gall
25	139	10.7	1450	1 MRF_CHICK	P36333 xenopus lae
26	137	10.6	1092	1 NCAL_HUMAN	P16573 ratius norv
27	136.5	10.5	519	1 ECTO_RAT	O08775 ratius norv
28	136.5	10.5	1343	1 VGR2_RAT	P10586 homo sapien
29	136.5	10.5	1897	1 PPRD_HUMAN	O13308 homo sapien
30	135.5	10.5	1070	1 PPRD_HUMAN	P40155 homo sapien
31	133.5	10.3	344	1 CEAL_HUMAN	O05793 mus musculu
32	133.5	10.3	3707	1 PGBM_MOUSE	P10721 homo sapien
33	132.5	10.2	976	1 KIT_HUMAN	

34	132	10.2	1377	1 NCAL_RAT	P97603 ratius norv
35	131.5	10.2	353	1 CEPU_CHICK	O90773 galus gall
36	131.5	10.2	526	1 NCAL_HUMAN	P13688 homo sapien
37	131	10.1	108	1 NCAL_XENLA	P16170 xenopus lae
38	130.5	10.1	298	1 NCAL_HUMAN	P57087 homo sapien
39	130.5	10.1	702	1 NCAL_HUMAN	P06721 homo sapien
40	130.5	10.1	864	1 TRKC_RAT	O03351 ratius norv
41	130	10.0	1367	1 VGR2_MOUSE	P25912 mus musculu
42	129.5	10.0	626	1 VGR2_MOUSE	P25912 mus musculu
43	129.5	10.0	822	1 FCRI_HUMAN	P1362 homo sapien
44	129.5	10.0	842	1 FCRI_MOUSE	P1692 mus musculu
45	129.5	10.0	822	1 FCRI_RAT	O04589 ratius norv

## ALIGNMENTS

RESULT 1  
ID NCAL\_BOVIN STANDARD: PRT: 853 AA.  
AC P31836;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)  
DE (NCAM-140)  
GN NCAM1 OR NCAM.  
OS Bos taurus (Bovine).  
OC Euryarchaea; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID:9913;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Brain cortex;  
RX MEDLINE=89378239; PubMed=2776887;  
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshayakov M.V.,  
RA Petukhova G.V., Rakitina T.V., Feshchenko E.A., Ishchenko K.A.,  
RA Marzoeva S.F., Chernova M.N., Dranyatsyna S.M.;  
RT "Calmodulin independent bovine brain adenylate cyclase. Amino acid  
RT sequence and nucleotide sequence of the corresponding cDNA.";  
RL FEBS Lett. 254:69-73(1989).  
RN [2]  
RP SEQUENCE OF 20-35.  
RX MEDLINE=86140120; PubMed=3512556;  
RA Rougon G., Marshak D.R.;  
RT "Structural and immunological characterization of the amino-terminal  
RT domain of mammalian neural cell adhesion molecules.";  
RL J. Biol. Chem. 261:3496-3401(1986).  
RN [3]  
RP IDENTIFICATION AS N-CAM  
RX MEDLINE=92111748; PubMed=1765159;  
RA Fremont R.T.;  
RT "A bovine brain cDNA purported to encode calmodulin-insensitive  
RT adenylate cyclase has extensive identity with neural cell adhesion  
RT molecules (N-CAMs)."  
RL FEBS Lett. 295:230-231(1991).  
CC - FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
CC NEURON-NEURON ADHESION. NEURITE FASCICULATION. GROWTH OF  
CC NEURITES, ETC.  
CC - SUPPLEMENTARY INFORMATION Type 1 membrane protein.  
CC - ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF  
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC - SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC - SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC - CAUTION: WAS ORIGINALLY (REF 1) THOUGHT TO BE A CALMODULIN-  
CC INDEPENDENT ADENYLATE CYCLASE.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way



FT	CARBOHYD	316	316	N-LINKED (GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	348	348	N-LINKED (GLCNAC. .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD	434	434	N-LINKED (GLCNAC. .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. .) <th>(POTENTIAL)</th>	(POTENTIAL)
FT	CARBOHYD	489	489	N-LINKED (GLCNAC. .) <th>(POTENTIAL)</th>	(POTENTIAL)
SO	SEQUENCE	858 AA;	94658 MW;	EA1A064AEAD050F6	CRC64;

Query Match	13.5%	Score: 175;	DR 1;	Length: 858.
Best Local Similarity	28.4%	Pred. No. 8.2e-07;		
Matches: 60;	Conservative: 31;	Mismatches: 88;	Indels: 32;	Gaps: 8;

QY	41	QASLIC-----AVQNTTREFELMYEEGRVGLKSGNKINS-----SSVCYS	82
Db	325	QVLTLCASDDPIPISTWRTSTRNLSSEKRSKMTPEKQELTGDHGVYRSHARVSSLTLLK	384

QY	83	SISNDNGISFTC-----RLGRQSVSVSVLYMTPTPLLSNDTQTYEESGNKLYCNK	138
Db	385	SIQYTDAG-EYITASNTIGD---SOSVLEYVQAPKLOPVAVYTWEGNQNITCEVY	440

QY	139	ANPQAMMYWKNSSLLDELEK-SRHOIQQRSEFSQISTIKVKEKPGNGTYSIAKSSIKTES	197
Db	441	AYSPATISWNRDQGLPSSNYSNIRKYNTPPSASTYLEVTPDSNDPCNCTAAYNRIGQES	500

QY	198	LDFELIVKDKTGVV---TEPIIACVYIF	224
Db	501	LEFTVQAD-TPSSPSIDREVEPYSIAGVQF	530

RESULT	3
DISCA_HUMAN	STANDARD; PRT: 2012 AA.
ID	060469; 060468;
AC	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Down syndrome cell adhesion molecule precursor (CHD2).
GN	DSGAM.
OS	Homo sapiens (Human).
OC	Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC	Mammalia, Eutheria, Plimates, Carnivora, Homiidae, Homo.
OX	NCBI.TaxId:9606;
RN	[1]
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC	TISSUE-Brain;
RX	MEDLINE-98087574; PubMed-9426258;
RA	Yamakawa K., Huot Y.-K., Haendel M. A., Hubert P., Chen X.-N.,
RA	Lyons G.E., Koenberg J.R.;
RT	DSGAM: a novel member of the immunoglobulin superfamily maps in a
RT	down syndrome region and is involved in the development of the
RT	nervous system.;
RL	Hum. Mol. Genet. 7:227-237(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	MEDLINE-20289799; PubMed-10830953;
RA	Hattori M., Fujiyama A., Taylor T.D., Matanabe H., Yada T.,
RA	Park H.-S., Toyoda A., Ishii K., Tokoki Y., Choi D.-K., Soeda E.,
RA	Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA	Meizel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA	Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA	Rosenfeld A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA	Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA	Mitsushima S., Shimizu N., Norisue K., Horisue J., Brandt P.,
RA	Scharfe M., Schoen O., Desario A., Reichelt J., Kaur G., Rieger H.,
RA	Ramser J., Beck A., Klages S., Hennig S., Rlesseman L., Dagand E.,
RA	Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA	Lehrach H., Reinhardt R., Yaspo M.-L.;
RT	"The DNA sequence of human chromosome 21.";
RL	Nature 405:311-319(2000).
RN	[3]
RP	FUNCTION.
RC	MEDLINE-20384934; PubMed-10925149;
RA	Agarwala K.L., Nakamura S., Tatusumi Y., Yamakawa K.;

RT	"Down syndrome cell adhesion molecule DSCAM mediates homophilic intercellular adhesion."
KL	Brain Res. Mol. Brain Res. 79:118-126(2000).
CC	-I- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN NERVOUS SYSTEM DEVELOPMENT.
CC	-I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE SHORT ISOFORM MAY BE SECRETED.
CC	-I- ALTERNATIVE PROMOTERS: 2 IDENTIFIED; A LONG PROMOTER <sup>1</sup> (SHOWN HERE) AND A SHORT FORM/CHDZ-42 ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	-I- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.
CC	-I- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC	-I- SIMILARITY: CONTAINS 10 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC	-I- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC	This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.ebi.ac.uk/announcements/">http://www.ebi.ac.uk/announcements/</a> or send an email to <a href="mailto:license@ebi.slb.ch">license@ebi.slb.ch</a> ).
DR	EMBL: AF023450; AAC17967.1; -
DR	EMBL: AF023449; AAC17966.1; -
DR	EMBL: AL161383; CAB90464.1; -
DR	EMBL: AL161382; CAB90436.1; -
DR	EMBL: AL161381; CAB90444.1; -
DR	HSSP: P40189; 1BOU.
DR	MIM: 602523; -
DR	InterPro: IPRO03961; FN_III.
DR	InterPro: IPRO03962; FNIII_repeat.
DR	InterPro: IPRO03006; Ig_MHC.
DR	InterPro: IPRO03598; Ig_C2.
DR	InterPro: IPRO03600; Ig_Like.
DR	Pfam: PF00041; fn3; 6.
DR	Pfam: PF00047; Ig; 9.
DR	PRINTS: PR00014; FNTPETIII.
DR	SMART: SM00060; FN3; 6.
LK	SMART: SM00410; IG_Like; 2.
LK	SMART: SM00408; ISC2; 7.
KM	Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat; Transmembrane; Alternative splicing.
KW	SIGNAL
FT	CHAIN 1 17
FT	FRAMES 19 2012
FT	DOMAIN 18 1595
FT	TRANSMEM 1596 1616
FT	DOMAIN 1617 2012
FT	DOMAIN 35 109
FT	DOMAIN 138 204
FT	DOMAIN 239 300
FT	DOMAIN 328 352
FT	DOMAIN 421 491
FT	DOMAIN 518 582
FT	DOMAIN 610 676
FT	DOMAIN 704 773
FT	DOMAIN 802 872
FT	DOMAIN 885 972
FT	DOMAIN 984 1076
FT	DOMAIN 1088 1177
FT	DOMAIN 1189 1273
FT	DOMAIN 1300 1366
FT	DOMAIN 1380 1463
FT	DOMAIN 1477 1562
FT	DISULFID 46 102
FT	DISULFID 145 197
FT	DISULFID 246 293
FT	DISULFID 335 385
FT	DISULFID 428 484
FT	DISULFID 525 575
FT	DISULFID 617 669
FT	DISULFID 711 766
FT	DISULFID 809 865





GN NCAM1 OR NCAM.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OK NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (N-CAM 120).  
 RP TISSUE-Skeletal muscle.  
 RX MEDLINE=89305258; PubMed=3253057;  
 RA Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,  
 RA Elton V., Moore S.E., Gordin C., Walsh F.S.,  
 RT \*Complete sequence and in vitro expression of a tissue-specific  
 RT phosphatidylinositol-linked N-CAM isoform from skeletal muscle.\*;  
 RL Development 104:165-173(1988).  
 RN [2]  
 RP SEQUENCE OF 491-655 FROM N.A. (N-CAM 120).  
 RP TISSUE-Skeletal muscle.  
 RX MEDLINE=87301755; PubMed=2887295;  
 RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,  
 RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;  
 RT \*Human muscle neural cell adhesion molecule (N-CAM): identification  
 RT of a muscle-specific sequence in the extracellular domain.\*;  
 RL Cell 50:1119-1130(1987).  
 RN [3]  
 RP SEQUENCE OF 491-655 FROM N.A. (SECRETED ISOFORM).  
 RX MEDLINE=89077552; PubMed=3203385;  
 RA Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,  
 RA Dickson G., Walsh F.S.;  
 RT \*Alternative splicing generates a secreted form of N-CAM in muscle  
 RT and brain.\*;  
 RL Cell 55:955-964(1988).  
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
 CC NEURITES, ETC.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF  
 CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD56 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd56.htm".  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X16841; CA34739.1; -;  
 DR EMBL: M17409; AA53912.1; -;  
 DR EMBL: M22094; AA59910.1; -;  
 DR EMBL: M22092; AA59911.1; -;  
 DR EMBL: M22091; AA59911.1; JOINED.  
 DR PIR: S07784; IJHUNG.  
 DR PIR: A31635; A31635.  
 DR MIM: 116930; -;  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003598; IG\_C2.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00047; ig; 5.  
 DR SMART: SM00060; FN3; 2.  
 DR SMART: SM00408; IGc2; 5.  
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; signal;  
 KW GPI-anchor; Alternative splicing.  
 FT SIGNAL 1 19  
 FT CHAIN 20 761 NEURAL CELL ADHESION MOLECULE.  
 FT DOMAIN 34 103 IG-LIKE C2-TYPE DOMAIN 1.  
 FT DOMAIN 132 196 IG-LIKE C2-TYPE DOMAIN 2.

FT DOMAIN 228 294 IG-LIKE C2-TYPE DOMAIN 3.  
 FT DOMAIN 322 392 IG-LIKE C2-TYPE DOMAIN 4.  
 FT DOMAIN 419 486 IG-LIKE C2-TYPE DOMAIN 5.  
 FT DOMAIN 518 595 FIBRONECTIN TYPE-III 1.  
 FT DOMAIN 660 727 FIBRONECTIN TYPE-III 2.  
 FT DISULFID 41 96 PROBABLE.  
 FT DISULFID 139 189 PROBABLE.  
 FT DISULFID 235 287 PROBABLE.  
 FT DISULFID 329 385 PROBABLE.  
 FT DISULFID 426 479 PROBABLE.  
 FT CARBOHYD 232 222 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 315 315 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 347 347 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 449 449 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 478 478 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSPLIC 635 655 GLPSPARKLEUCMEQENSLIKV -> NLAQNHCCNHQNGLH  
 FT NALMK (IN SECRETED ISOFORM).  
 SQ SEQUENCE 761 AA; 83770 MM; F0CAD3292D7AB67E CRC64;  
 Query Match 12.8%; Score 166; DB 1; Length 761;  
 Best Local Similarity 28.0%; Pred. No. 3.9e-06;  
 Matches 58; Conservative 35; Mismatches 80; Indels 34; Gaps 9;  
 QY 41 QASLIC-----AVQNHTRDEELMYREGGRVLDKSGKINSSTCVASISSE 86  
 DB 324 QVTLCEASGDPISPTWRTSTRNISSEKTL-----DGHVYVSHARV--SSLTLSIOY 377  
 QY 87 NDNGISFPC---KRGKRVSVSVLVNVIHPELISGNDVQIVVEESNVKLVNKNANV 142  
 DB 378 TIVAT-PTCTGASNTIQQ---SGSMYLFVOYAFKLDGPVAVYVTEGNOVITTEVAYVS 433  
 QY 143 AQMMYKNSLLEDEK-SPHQIQVLSSEPSULITVERPQVNTYSVIAKNSIKFSTH 201  
 DB 434 ATISFRDGLSPSSNYSNIKIYVTPASYLEVTPDSENDENGNCTAVNRIGQESLEFI 493  
 QY 202 LIVKKTGVG---IEPIIACVVF 224  
 DB 494 LVQAV-LVSSPSLDVEPYSTAVQVF 519  
 RESULT 6  
 NCAM\_HUMAN STANDARD; PRT; 848 AA.  
 AC P13591; Q16180; Q15829;  
 UT 01-JAN-1990 (Rel. 13, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)  
 DE (NCAM-140) (CD56 antigen).  
 GN NCAM1 OR NCAM.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OK NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94356433; PubMed=8075973;  
 RA Saito S., Iano Y., Iachilava I., Hayashi S., Kishimoto T., Kawase I.,  
 RA Larter L.W., Chang C., Azuma M., Kuitenberg J.J., Hemperly J.J.,  
 RA Phillips J.H.;  
 RT \*Molecular and functional analysis of human natural killer cell-  
 RT associated neural cell adhesion molecule (N-CAM/CD56)\*;  
 RL J. Immunol. 146:4421-4426(1991).  
 RN [3]  
 RP SEQUENCE OF 491-848 FROM N.A.

[illegible]

```

Query Match      12.8%; Score 166; DB 1; Length 848;
Best Local Similarity    28.0%; Pred. No. 4,4e-06;
Matches   56; Conservative   35; Mismatches   80; Indels   34; Gaps

```

CY	41	QASTIT-----AVGNTHPTPLTWPPHKKIKKINKINSVSVSLSE	86
DB	324	CVLLTFLASDPLPSLWRLSIRKLSIRSKKLT----LADVVSSAKV--SSLTKSTGY	377
OY	87	NINISFTG---RIKROGSVSVALWTBPLISGDFQVEESNKVLGCNKNRQ	142
Lb	378	LEAS-ELICLASNTIGD---SGSYMLELVYAPFLLGVPAVTWRNSNVNITTEHPATFS	433
CY	143	KCKMRTNNESTILGTF--SNNGTGGLSPSSGCT--LTVKGRMLIVS--IAVSSPIELDPH	201
DB	434	AIIISWEKDLDEPSSNNISKIYNIPSAVILEVFGSDENFDSNM--LAVRKGQVSELEF	493
OY	202	LIVKUNLVSVV-----LEPIAACVVIF	224
DB	494	LYOADTPRSSPIDOYEFPSSAQOF	519

RESULT	7	
NCAL	CHICK	STANDARD:
ID	NCAL	PRT: 1091 AA.
AC	P13590;	090919; 090918;
DT	01-JAN-1990 (Rel. 13, Created)	
DT	15 JUL 1998 (Rel. 36, Last sequence update)	
DT	01 MAR 2002 (Rel. 41, Last annotation update)	
DE	Neural cell adhesion molecule, large isoform precursor (N-CAM 180)	
DE	[contains: N-CAM 140].	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae	
OC	Gallus.	
OX	NCBI_TaxID=9041;	
LN	11	
RP	SEQUENCE OF 1-175 FROM N.A. AND PARTIAL SEQUENCE.	
RX	MEDLINE=87206190; PubMed=3576199;	
RA	Cunningham B.A., Hemperly J.J., Murray H.A., Prediger R.A.,	
RA	Blackenbury R., Edelman G.M.;	
RT	"Neural cell adhesion molecule: structure, immunoglobulin-like	
RT	domains, cell surface mediators, and alternative RNA splicing.";	
RL	Science 236:799-806(1987).	
RN	12	
RP	SEQUENCE OF 128-1091 FROM N.A. AND PARTIAL SEQUENCE.	
RX	MEDLINE=86205089; PubMed=3458261;	
RA	Hemperly J.J., Murray H.A., Edelman G.M., Cunningham B.A.;	
RT	"Sequence of a cDNA clone encoding the polypeptide acid-rich and	
RT	cytoplasmic domains of the neural cell adhesion molecule N-CAM.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 83:3037-3041(1986).	
RN	13	
RP	SEQUENCE FROM N.A. (726 AA FROM).	
RX	MEDLINE=87092340; PubMed=3467341;	
RA	Hemperly J.J., Edelman G.M., Cunningham B.A.;	
RT	"cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a	
RT	membrane spanning region, consistent with evidence for membrane	
RT	attachment via a phosphatidylinositol intermediate.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 83:9822-9826(1986).	
RN	14	
RP	SEQUENCE OF 810-1064 FROM N.A.	
RX	MEDLINE=87011934; PubMed=3771645;	
RA	Merrif B.A., Owens G.C., Prediger E.A., Crossin K.L.,	
PA	Cunningham B.A., Edelman G.M.;	
RT	"Cell surface modulation of the neural cell adhesion molecule	
RT	resulting from alternative mRNA splicing in a tissue-specific	
RT	developmental sequence.";	
RL	J. Cell Biol. 103:1431-1439(1986).	
RN	15	
RP	SEQUENCE OF 1-17 FROM N.A.	
RX	MEDLINE=93122797; PubMed=1478668;	
RA	Colwell G., Li B., Forrest D., Brackenbury R.;	
RT	"Conserved regulatory elements in the promoter region of the N-CAM	

RT	gene."
RL	Genomics 14:875-882(1992).
RN	[6]
RP	SEQUENCE OF 1-17 FROM N.A.
RC	STRAIN-WHITE LEIGHORN; TISSUE-Erythrocyte;
RA	Sasner M., Covault J.;
RL	Submitted (Aug-1993) to the EMBL/GenBank/DDIY databases.
CC	- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF NEURITES, ETC.
CC	- SUBCELLULAR LOCATION: Type I membrane protein.
CC	- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	- SIMILARITY: BELONGS TO THE IMMUGLOBULIN SUPERFAMILY.
CC	- SIMILARITY: CONTAINS 5 IMMUGLOBULIN-LIKE C2-TYPE DOMAINS.
CC	- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC	-----
CC	This SWISS-Prot entry is copyright © It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; M15861; AAB59958.1; JOINED.
DR	EMBL; M15860; AAB59958.1; JOINED.
DR	EMBL; M15022; AAB59958.1; JOINED.
DR	EMBL; M15923; AAB59958.1; JOINED.
DR	EMBL; M15924; AAB59958.1; JOINED.
DR	EMBL; M21178; AAB59958.1; JOINED.
DR	EMBL; M21179; AAB59958.1; JOINED.
DR	EMBL; M21180; AAB59958.1; JOINED.
DR	EMBL; M15929; AAB59958.1; JOINED.
DR	EMBL; M15929; AAB59958.1; JOINED.
DR	EMBL; M15935; AAB59958.1; JOINED.
DR	EMBL; M15937; AAB59958.1; JOINED.
DR	EMBL; M15938; AAB59958.1; JOINED.
DR	EMBL; M15939; AAB59958.1; JOINED.
DR	EMBL; M15861; AAB59958.1; JOINED.
DR	EMBL; M15860; AAB59958.1; JOINED.
DR	EMBL; M15922; AAB59958.1; JOINED.
DR	EMBL; M15923; AAB59958.1; JOINED.
DR	EMBL; M15924; AAB59958.1; JOINED.
DR	EMBL; M21178; AAB59958.1; JOINED.
DR	EMBL; M21179; AAB59958.1; JOINED.
DR	EMBL; M21180; AAB59958.1; JOINED.
DR	EMBL; M15929; AAB59958.1; JOINED.
DR	EMBL; M15930; AAB59958.1; JOINED.
DR	EMBL; M15931; AAB59958.1; JOINED.
DR	EMBL; M15932; AAB59958.1; JOINED.
DR	EMBL; M15933; AAB59958.1; JOINED.
DR	EMBL; M15934; AAB59958.1; JOINED.
DR	EMBL; M15935; AAB59958.1; JOINED.
DR	EMBL; M15936; AAB59958.1; JOINED.
DR	EMBL; M15937; AAB59958.1; JOINED.
DR	EMBL; M15938; AAB59958.1; JOINED.
DR	EMBL; M15939; AAB59958.1; JOINED.
DR	EMBL; M15860; AAB59958.1; JOINED.
DR	EMBL; M15922; AAB59958.1; JOINED.
DR	EMBL; M15923; AAB59958.1; JOINED.
DR	EMBL; M15924; AAB59958.1; JOINED.
DR	EMBL; M21178; AAB59958.1; JOINED.
DR	EMBL; M21179; AAB59958.1; JOINED.
DR	EMBL; M21180; AAB59958.1; JOINED.
DR	EMBL; M15929; AAB59958.1; JOINED.
DR	EMBL; M15930; AAB59958.1; JOINED.
DR	EMBL; M15931; AAB59958.1; JOINED.
DR	EMBL; M15932; AAB59958.1; JOINED.
DR	EMBL; M15933; AAB59958.1; JOINED.
DR	EMBL; M15934; AAB59958.1; JOINED.
DR	EMBL; M15935; AAB59958.1; JOINED.
DR	EMBL; M15936; AAB59958.1; JOINED.
DR	EMBL; M15937; AAB59958.1; JOINED.
DR	EMBL; M15938; AAB59958.1; JOINED.
DR	EMBL; M15939; AAB59958.1; JOINED.
DR	EMBL; M15860; AAB59958.1; JOINED.
DR	EMBL; M15922; AAB59958.1; JOINED.
DR	EMBL; M15923; AAB59958.1; JOINED.
DR	EMBL; M15924; AAB59958.1; JOINED.
DR	EMBL; M21178; AAB59958.1; JOINED.
DR	EMBL; M21179; AAB59958.1; JOINED.
DR	EMBL; M21180; AAB59958.1; JOINED.
DR	EMBL; M15929; AAB59958.1; JOINED.
DR	EMBL; M15930; AAB59958.1; JOINED.
DR	EMBL; M15931; AAB59958.1; JOINED.
DR	EMBL; M15932; AAB59958.1; JOINED.
DR	EMBL; M15933; AAB59958.1; JOINED.
DR	EMBL; M15934; AAB59958.1; JOINED.
DR	EMBL; M15935; AAB59958.1; JOINED.
DR	EMBL; M15936; AAB59958.1; JOINED.
DR	EMBL; M15937; AAB59958.1; JOINED.
DR	EMBL; M15938; AAB59958.1; JOINED.
DR	EMBL; M15939; AAB59958.1; JOINED.
DR	EMBL; M15860; AAB59958.1; JOINED.
DR	EMBL; M15922; AAB59958.1; JOINED.
DR	EMBL; M15923; AAB59958.1; JOINED.
DR	EMBL; M15924; AAB59958.1; JOINED.
DR	EMBL; M21178; AAB59958.1; JOINED.
DR	EMBL; M21179; AAB59958.1; JOINED.
DR	EMBL; M21180; AAB59958.1; JOINED.
DR	EMBL; M15929; AAB59958.1; JOINED.
DR	EMBL; M15930; AAB59958.1; JOINED.
DR	EMBL; M15931; AAB59958.1; JOINED.
DR	EMBL; M15932; AAB59958.1; JOINED.
DR	EMBL; M15933; AAB59958.1; JOINED.
DR	EMBL; M15934; AAB59958.1; JOINED.
DR	EMBL; M15935; AAB59958.1; JOINED.
DR	EMBL; M15936; AAB59958.1; JOINED.
DR	EMBL; M15937; AAB59958.1; JOINED.
DR	EMBL; M15938; AAB59958.1; JOINED.
DR	EMBL; M15939; AAB59958.1; JOINED.
DR	EMBL; M15860; AAB59958.1; JOINED.
DR	EMBL; M15922; AAB59958.1; JOINED.
DR	EMBL; M15923; AAB59958.1; JOINED.
DR	EMBL; M15924; AAB59958.1; JOINED.
DR	EMBL

DR	EMBL:	L15934:	AAB59957.1:	JOINED.
DR	EMBL:	L29437:	AAB59957.1:	JOINED.
DR	EMRL:	M15935:	AAB59957.1:	JOINED.
DR	EMBL:	X04479:	CAB51638.1:	-
DR	EMBL:	X70342:	CAA49807.1:	-
DR	EMBL:	Z12128:	CAA78113.1:	-
DR	PIR:	A43613:	IJCENL.	-
DR	HSSP:	O13740:	IKUC.	-
DR	InterPro:	IPRO033961:	FN_III.	-
DR	InterPro:	IPRO03006:	Ig_HRC.	-
DR	InterPro:	IPRO03598:	Ig_C2.	-
DR	InterPro:	IPRO03600:	Ig_Like.	-
DR	Pfam:	PF00041:	fn3: 2.	-
DR	Pfam:	PF00047:	fn3: 5.	-
DR	SMART:	SMO0060:	FM3: 2.	-
DR	SMART:	SMO0410:	IC_Like: 1.	-
DR	SMART:	SMO0408:	JGc2: 4.	-
KW	Cell adhesion:	Glycoprotein:	Transmembrane; Repeat;	
KM	Immunoglobulin domain:	Alternative splicing:	Signal.	
FT	SIGNAL	1	19	
FT	CHAIN	20	1091	
FT	DOMAIN	20	711	
FT	TRANSMEM	712	729	
FT	DOMAIN	730	1091	
FT	DOMAIN	34	103	
FT	DOMAIN	132	196	
FT	DOMAIN	228	294	
FT	DOMAIN	322	392	
FT	DOMAIN	419	486	
FT	DOMAIN	518	595	
FT	DOMAIN	624	692	
FT	DOMAIN	152	156	
FT	DOMAIN	161	165	
FT	DISULFID	41	96	
FT	DISULFID	139	189	
FT	DISULFID	235	287	
FT	DISULFID	329	385	
FT	DISULFID	426	479	
FT	CARBOHYD	222	222	
FT	CARBOHYD	315	315	
FT	CARBOHYD	347	347	
FT	CARBOHYD	423	423	
FT	CARBOHYD	445	445	
FT	CARBOHYD	478	478	
FT	VAPSPIT	702	726	
FT	VAPSPIT	727	1091	
FT	VAPSPIT	771	809	
FT	VAPSPIT	810	1070	
ST	SEQUENCE	1091 AA:	117415 MM:	B1J156/U0ZDFB8BU CXC64;

Query Match 12.4%; Score 160.5; DB 1: length 1091;  
Best Local Similarity 26.9%; Pred. No. 1.7e-05;  
Matches 60; Conservative 37; Mismatches 87; Indels 39; Gaps 10;

```

QY      25 VNKTENTVILDTTPSGASLIC-----AQNCTREBELLMKREGCKVDLS 70
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB      313 VENTAMELED-----QITLTCEASGDPIPSITWKTSIRNISNEKPL---DCRIYVR 363
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY      71 GNKNSSSVCSISINDNGISFTC-----RLGRDSVSYSVLAWTPPLSLSDNDOTVE 126
          ::||::||::||::||::||::||::||::||::||::||::||::||::||
DB      364 HAKV--SLLTLKIQLIUTUNG-EVYCIVASNTIGUD---SQAMYLEVUAFKLQGPVAVTM 417
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
QY     127 EGSNVKLYCVANKNPOAOAMMTKNSSLDDLEK-SRHQIQUTSESPQLSTRVKRPONGTY 185
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
LB      418 EGNNVNITLCVFVFPFAVISWEKDCGLTLESSNYMKIKIYNTPSAISLEVTPUSESDPGNY 477
          ||| | | | | | | | | | | | | | | | | | | | | | | | | |
CY     186 SCIAKSKTESLDPHLIYKDKTVCP----IEPLIACVYIF 224
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      478 NCTAVNRHIGESSEFILOYAD-TPSSPFIIDKEEPISTARVEF 519
          :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 9  
ID NCAL\_MOUSE STANDARD: PRT: 725 AA.  
AC P13594; Q61950;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Neural cell adhesion molecule, phosphatidylinositol-linked isoform  
OR NCAM1 OR NCAM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RX MEDLINE=87246524; PubMed=3595563;  
RA Barthols D., Santoni M.-C., Wille M., Ruppert C., Calix J.-C.,  
RT Birbach M.-H., Fontecilla-Camps J.-C., Cordis G.;  
KT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for  
a Mr 79,000 polypeptide without a membrane-spanning region."  
RI PubMed J. 6:907-914(1987).  
RN [2]  
RP SEQUENCE OF 20 700 FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=89251564; PubMed=2721486;  
RA Santoni M.-H., Barthols D., Vopper G., Boud A., Cordis G., Wille M.;  
RT "Differential exon usage involving an unusual splicing mechanism  
generates at least eight types of NCAM cDNA in mouse brain."  
RI PubMed J. 8:185-192(1989).  
RN [3]  
RP SEQUENCE OF 642-725 FROM N.A.  
RX MEDLINE=86281628; PubMed=3395544;  
RA Barnes J.A., Calix J.C., Steinmetz M., Cordis G.;  
RT "Differential splicing and alternative polyadenylation generates  
distinct NCAM transcripts and proteins in the mouse."  
RI PubMed J. 7:625-642(1988).  
RN [4]  
RP SEQUENCE OF 20-76.  
RX MEDLINE=86140120; PubMed=3512556;  
RA Rougon G., Marthok D.R.;  
RT "Structural and immunological characterization of the amino-terminal  
domain of mammalian neural cell adhesion molecules."  
RI J. Biol. Chem. 261:3396-3401(1986).  
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
NEURITES, ETC.  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -1- ALTERNATIVE PRODUCTS: The different tissue-specific forms of  
N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
CC This SwissProt entry is a duplicate. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.ebi.ac.uk/submit/submit.html>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR Pfam: PF00047; Iq: 5.  
DR SMART: SM00060; FN3: 2.  
DR SMART: SM00408; IG2: 5.  
KW Cell adhesion; Glycoprotein; Repeat; Alternative splicing;  
KW Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor;  
FT SIGNAL 1 19  
FT CHAIN 20 725  
FT  
FT 1 LAMIN 34 103  
FT DOMAIN 132 196  
FT DOMAIN 228 295  
FT DOMAIN 323 393  
FT DOMAIN 420 487  
FT DOMAIN 519 596  
FT DOMAIN 625 692  
FT DOMAIN 152 156  
FT DOMAIN 161 165  
FT DOMAIN 161 165  
FT DISULFID 41 96  
FT DISULFID 139 189  
FT DISULFID 235 288  
FT DISULFID 340 386  
FT DISULFID 427 480  
FT CARBOHYD 222 222  
FT CARBOHYD 316 316  
FT CARBOHYD 348 348  
FT CARBOHYD 424 424  
FT CARBOHYD 450 450  
FT CARBOHYD 479 479  
FT CONFLICT 261 268  
FT CONFLICT 273 273  
FT CONFLICT 354 355  
FT CONFLICT 549 549  
FT CONFLICT 572 572  
FT CONFLICT 575 575  
FT CONFLICT 589 594  
FT CONFLICT 690 692  
FT CONFLICT 657 657  
SQ SEQUENCE 125 AA; 80496 MW; G2AEH844b16B2F CRC64;  
Query Match 12.38; Score 159; DB 1; Length 725;  
Best Local Similarity 27.58; Pred No. 1,4e-05;  
Matches 57; Conservative 35; Mismatches 81; Indels 34; Gaps 9;  
QY 41 CASLIK-----AVONTPEELIMREESPVLSKANKNSVCSYSIE 66  
DB 325 QVLTCEASGDPISITWTSFRNISSEQDL---DSHMVVRSHAV--SSLTLSIOY 178  
QY 87 NONGISFIC---KQKQSVSVSVINVI+PPILSONM+P+VHSGSNVKNVNVANPQ 142  
DB 476 PMAE+TVM+TASNTDGL+---S+SLDIF+P+YAP+P+L+G+VAVY+TW+DQVNT+LE+VAVPS 414  
QY 144 A+MMWY+N+S+L+L+L+K+---S+H+L+Q+L+S+L+S+L+K+V+K+P+N+I+S+L+V+S+L+K+P+S+L+H 201  
DB 435 AI+SM+K+D+G+L+P+S+S+N+I+K+Y+N+P+S+V+S+L+V+E+V+D+S+N+I+G+N+I+V+N+I+G+S+S+L+P 404  
QY 222 L+V+L+K+V+G+V+---L+P+T+I+M+G+V+T+ 224  
DB 495 LVQAD+TPSSPSTDEVPEPVSIAVOVF 520  
RESULT 9  
ID NCAL\_MOUSE STANDARD: PRT: 1115 AA.  
AC P13594; Q61949;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Neural cell adhesion molecule, large isoform precursor (N-CAM 180)  
OR (NCAM-180) [contains: N-CAM 140 (NCAM-140)].  
GN NCAM1 OR NCAM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (N-CAM 180).  
 RC STRAIN-C57BL/6;  
 RX MEDLINE=87246524; PubMed=3595563;  
 RA Barthels D., Santoni M.J., Wille W., Ruppert C., Chai J.C.,  
 RT Hirsch M.R., Fontecilla-Camps J.C., Goridis C.;  
 RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for  
 RT a Mr 79,000 polypeptide without a membrane-spanning region.";  
 RL EMBL J. 6:907-914(1987)  
 RN [2]  
 RP SEQUENCE OF 529-1115 FROM N.A. (N-CAM 140).  
 RC STRAIN-C57BL/6;  
 RX MEDLINE=86067687; PubMed=3684567;  
 RA Santoni M.-J., Barthels D., Barbas J.A., Hirsch M.-R., Steinmetz M.,  
 RT Goridis C., Wille W.;  
 RT "Analysis of cDNA clones that code for the transmembrane forms of the  
 RT mouse neural cell adhesion molecule (NCAM) and are generated by  
 RT alternative RNA splicing.";  
 RL Nucleic Acids Res. 15:8621-8641(1987).  
 RN [3]  
 RP SQUINCE OF 642-1115 FROM N.A. (N-CAM 180).  
 RC MEDLINE=88283628; PubMed=3396534;  
 RA Barbas J.A., Chai J.C., Steinmetz M., Goridis C.;  
 RT "Differential splicing and alternative polyadenylation generates  
 RT distinct NCAM transcripts and proteins in the mouse.";  
 RL EMBL J. 7:625-632(1988).  
 RN [4]  
 RP SEQUENCE OF 804-1081 FROM N.A. (N-CAM 180).  
 RC STRAIN-C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=88247737; PubMed=2454455;  
 RA Barthels D., Vopper G., Wille W.;  
 RT "NCAM-180, the large isoform of the neural cell adhesion molecule of  
 RT the mouse, is encoded by an alternatively spliced transcript.";  
 RL Nucleic Acids Res. 16:4217-4225(1988)  
 RN [5]  
 RP SEQUENCE OF 702-1115 FROM N.A. (N-CAM 140).  
 RC STRAIN-C57BL/6; TISSUE=Brain;  
 RX MEDLINE=89251563; PubMed=2721486;  
 RA Santoni M.J., Barthels D., Vopper G., Roneg A., Goridis C., Wille W.;  
 RT "Differential exon usage involving an unusual splicing mechanism  
 RT generates at least eight types of NCAM cDNA in mouse brain.";  
 RL EMBL J. 8:385-392(1989).  
 RN [6]  
 RP SEQUENCE OF 20-36.  
 RX MEDLINE=86140120; PubMed=3512556;  
 RA Rougon G., Marshak D.R.;  
 RT "Structural and immunological characterization of the amino-terminal  
 RT domain of mammalian neural cell adhesion molecules.";  
 RL J. Biol. Chem. 261:3396-3401(1986).  
 CC - FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
 CC NEURITES, ETC.  
 CC - SUPRACELLULAR LOCATION: TYPE I membrane protein.  
 CC - ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF  
 CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC - SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC - SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC - SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL, on condition  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X07200; CAA30177.1; -;  
 DR EMBL: Y00051; -; NOT ANNOTATED\_CDS.  
 DR EMBL: X06328; CAA29641.1; -;  
 DR EMBL: X07195; CAA30173.1; -;

DR EMBL: X07244; CAA30230.1; -;  
 DR EMBL: X15051; CAA33150.1; -;  
 DR EMBL: X15052; CAA33151.1; -;  
 DR PIR: A29673; IJMSNL.  
 DR MGD: MGI:97281; NCAM.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00047; Ig; 5.  
 DR SMART: SM00060; FN3; 2.  
 DR SMART: SM00408; IGC2; 5.  
 KM Cell adhesion; glycoprotein; Transmembrane; Repeat;  
 KM Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.  
 FT SIGNAL 1 19  
 FT CHAIN 20 115  
 FT  
 FT DOMAIN 20 711  
 FT TRANSMEM 712 729  
 FT DOMAIN 730 1115  
 FT DOMAIN 34 103  
 FT DOMAIN 132 196  
 FT DOMAIN 228 295  
 FT DOMAIN 323 393  
 FT DOMAIN 420 487  
 FT DOMAIN 519 596  
 FT DOMAIN 625 692  
 FT DOMAIN 161 165  
 FT DISULF 41 96  
 FT DISULF 139 189  
 FT DISULF 235 288  
 FT DISULF 330 386  
 FT DISULF 427 480  
 FT CARBOHYD 222 222  
 FT CARBOHYD 318 318  
 FT CARBOHYD 348 348  
 FT CARBOHYD 424 424  
 FT CARBOHYD 450 450  
 FT CARBOHYD 479 479  
 FT VARPSP 810 1076  
 SV SEQUENCE 1115 AA; 119351 MW; 2093DD474CEBCAF CRC64;  
 Query Match 12.3%; Score 159; DB 1; Length 1115;  
 Best Local Similarity 27.5%; Pred. No. 2,4e-05;  
 Matches 57; Conservative 35; Mismatches 81; Indels 34; Gaps 9;  
 QY 41 VASLLC-----AVUNTFREELLMREELKVDLSGKINSNVCSVSI 86  
 DB 325 VYLTCESAGDPPIETWTETRNISSEQDL---DSHMVYRSHAV--SLITLKI 378  
 QY 87 NUNGLSTIC---ELSLKVSASVSVLVNTPPLLSNMFEFTVEFSNVFVNVANPQ 142  
 DB 379 PLAG-EYMTASNTIGQD---SQSIDLEFYAPKLGVAAYTEGQVNTCEVFAVPS 434  
 QY 143 AQMMVYKKNLSLLLEK-SHNUIGQISBSFQSLIKVHLMNGVTSCLAKSTKTESIDF 201  
 DB 435 AIIISWEDQQLLPSSNYSNKLNTNPASYLEVTPDSENDGNTCTAVNPIGQESLEFI 494  
 QY 202 LTVKDTGVP----IEPIIACVVI 224  
 DB 495 LVQAD-TFSSTLQVVERYSIAVUF 520  
 LAR\_FROME STANJAKU; PRT: 2029 AA.  
 AC P16621;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase lar precursor (EC 3.1.3.48) (Protein-



CC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
 OK NCBI\_TaxID=9606;  
 RN PIR: A54100; A54100.  
 RN PIR: A40098; A40098.  
 RN PIR: A38442; A38442.  
 RN BSS: F56276; 1TLK.  
 DR MIM: 120470; -  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003962; FN\_III\_repeat.  
 DR InterPro: IPR003006; 19\_MHC.  
 DR InterPro: IPR003598; 19\_C2.  
 DR InterPro: IPR003600; 19\_Like.  
 DR Pfam: PF00041; fn3; 6.  
 DR Pfam: PF00047; 19; 4.  
 DR PRINTS: PRO0014; FNTPELII.  
 DR SMART: SM00060; FN3; 5.  
 DR SMART: SM00410; IG\_Like; 2.  
 DR SMART: SM00408; IGC2; 3.  
 KW Glycoprotein, Immunoglobulin domain, Transmembrane, Signal;  
 KW Repeat, Anti-viral, Immune, Disulfide mutation, Polymorphism.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1447  
 FT DOMAIN 26 1097  
 FT TRANSMEM 1098 1122  
 FT DOMAIN 1123 1447  
 FT DOMAIN 54 124  
 FT DOMAIN 154 219  
 FT DOMAIN 254 317  
 FT DOMAIN 345 407  
 FT DOMAIN 426 522  
 FT DOMAIN 525 618  
 FT DOMAIN 619 716  
 FT DOMAIN 722 816  
 FT DOMAIN 840 940  
 FT DOMAIN 941 1042  
 FT DISULFD 61 117  
 FT DISULFD 161 212  
 FT DISULFD 261 310  
 FT DISULFD 352 400  
 FT CARBOHYD 94 299  
 FT CARBOHYD 299 318  
 FT CARBOHYD 318 318  
 FT CARBOHYD 478 478  
 FT CARBOHYD 628 628  
 FT CARBOHYD 702 702  
 FT VARIANT 168 168  
 FT VARIANT 201 201  
 FT VARIANT 1375 1375  
 FT CONFLICT 138 138  
 FT CONFLICT 233 329  
 FT CONFLICT 421 421  
 SV SEQUENCE 1447 AA; 158456 MW; 4A8612766ED0471F CRC64;  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

UR EMBL: M63698; AAA52181.1; -  
 DR PIR: A54100; A54100.  
 DR PIR: A40098; A40098.  
 DR PIR: A38442; A38442.  
 DR BSS: F56276; 1TLK.  
 DR MIM: 120470; -  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003962; FN\_III\_repeat.  
 DR InterPro: IPR003006; 19\_MHC.  
 DR InterPro: IPR003598; 19\_C2.  
 DR InterPro: IPR003600; 19\_Like.  
 DR Pfam: PF00041; fn3; 6.  
 DR Pfam: PF00047; 19; 4.  
 DR PRINTS: PRO0014; FNTPELII.  
 DR SMART: SM00060; FN3; 5.  
 DR SMART: SM00410; IG\_Like; 2.  
 DR SMART: SM00408; IGC2; 3.  
 KW Glycoprotein, Immunoglobulin domain, Transmembrane, Signal;  
 KW Repeat, Anti-viral, Immune, Disulfide mutation, Polymorphism.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1447  
 FT DOMAIN 26 1097  
 FT TRANSMEM 1098 1122  
 FT DOMAIN 1123 1447  
 FT DOMAIN 54 124  
 FT DOMAIN 154 219  
 FT DOMAIN 254 317  
 FT DOMAIN 345 407  
 FT DOMAIN 426 522  
 FT DOMAIN 525 618  
 FT DOMAIN 619 716  
 FT DOMAIN 722 816  
 FT DOMAIN 840 940  
 FT DOMAIN 941 1042  
 FT DISULFD 61 117  
 FT DISULFD 161 212  
 FT DISULFD 261 310  
 FT DISULFD 352 400  
 FT CARBOHYD 94 299  
 FT CARBOHYD 299 318  
 FT CARBOHYD 318 318  
 FT CARBOHYD 478 478  
 FT CARBOHYD 628 628  
 FT CARBOHYD 702 702  
 FT VARIANT 168 168  
 FT VARIANT 201 201  
 FT VARIANT 1375 1375  
 FT CONFLICT 138 138  
 FT CONFLICT 233 329  
 FT CONFLICT 421 421  
 SV SEQUENCE 1447 AA; 158456 MW; 4A8612766ED0471F CRC64;  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

Query Match 11.5%: Score 149; DB 1; Length 1447;  
 Best local similarity 25.8%;  
 Matches 51; Conservative 37; Mismatches 86; Indels 24; Gaps 9;  
 DB 11 ILFPEHREHSSVLYVNGTENTYLLDTTGSQASLCAVQNTREELMWRHFGVPLNS 70  
 DB 240 LKFLK-PSNVVALEK-----DAVLECCVSGVP-PPSTWLRGEEVYGLRS 284  
 QY 71 G--KINSSVCVSSISENDGISTFCRLG-RDOSSVSVVLYNTPPP-LISGNDQOTVE 126  
 DB 285 KRYSLGGSNLLISVNTDDSGM-VTCVVTYKNENISASALTLVLPWPLNHSNLYAY 343  
 QY 127 EGSNFKLVGVKANPQAOAMMYKNNSLDEKSRHIOQGTSEFQSLSTVEKPDNCTYS 186  
 DB 344 ESMDFIEFCTVSGKRPVPIVNMKNKGDV-IPSDVPIVIGGS---NLPILVASDECFYQ 399  
 QY 187 CLAKSLKTESLDPLHIV 204

db 400 CVAENAGNAGNAGTSAQIV 417

```

RESULT 12
PRTD_HUMAN STANFORD: PRT: 1913 AA.
AC P24468;
DT 01-NOV-1991 (rel. 20, last sequence update)
DT 01-OCT-1996 (rel. 34, last annotation update)
DT 01-MAR-2002 (rel. 41, last annotation update)
DE Protein-tyrosine phosphatase delta precursor (EC 3.1.3.48) (R-PTP-
delta).
GN PTPRD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
OX MCh_TaxId=9606;
RN 111
RP SOURCE FROM N.A., AND MUTAGENESIS OF AKO-1178.
RX MEDLINE-95204468; PubMed-7896816;
RA Follis R., Krueger N.X., Serra-Pons C., Saito H., Saito M.;
RT Molecular characterization of the human transmembrane protein-
RT tyrosine phosphatase delta. Evidence for tissue-specific expression of
RT alternative human transmembrane protein-tyrosine phosphatase delta
RT isoforms.*
RT J. Biol. Chem. 270:6722-6728(1995).
RN 121
RP SOURCE OF 390-1912 FROM N.A.
KC TISSUE-PLACENTA;
RX MEDLINE-91006019; PubMed-2170109;
RA Krueger N.X., Saito M., Saito H.;
RT Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases.*
RT J. Biol. Chem. 270:3241-3252(1995).
CC -1 CATALYTIC ACTIVITY: protein-tyrosine phosphate + H2O -> protein
CC tyrosine + phosphate.
CC -1 SUBCELLULAR LOCATION: Type I membrane protein.
CC -1 ALTERNATIVE PRODUCTS: TYPED DIFFERENT ISOPROPS AAF FOUND IN DIFFERENT
CC TISSUES BUT TO ALTERNATIVE SPLICING.
CC -1 PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
CC FROM THE TRANSMEMBRANE SEGMENT.
CC -1 SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1 SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1 SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC
CC This Swiss-Prot entry is a copy of the EMBL database. It is provided through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/annouce/).
CC
CC -----
CC EMBL: L36929; AAC14749.1;
CC EMBL: X54113; CAA38068.1;
CC EMBL: S12652; S12652.
CC HSSP: P18652; TYRO.
CC MIM: 601598;
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR003962; FNIII_repeat.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR003598; Ig_C2.
CC InterPro: IPR003877; TYR_phosphatase.
CC InterPro: IPR002442; TYR_prot_phphatase.
CC Pfam: PF00041; Ig_3.
CC Pfam: PF00047; Ig_3.
CC PRINTS: PR00014; FNTHP111.
CC PRINTS: PR00700; PTPPHPTASE.
CC SMART: SM00060; FN3; 8.
CC SMART: SM00408; IgC2; 3.
CC SMART: SM00194; PTPC; 2.

```

```

DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE: PS50956; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS50955; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane; Repeat;
KW Immunoglobulin domain; Alternative splicing.
FT SIGNAL: 1 20
FT CHAIN: 31 1912
FT DOMAIN: 21 1265
FT TRANSMEM: 1266 1290
FT DOMAIN: 1291 1912
FT DOMAIN: 23 115
FT DOMAIN: 118 225
FT DOMAIN: 232 318
FT DOMAIN: 320 414
FT DOMAIN: 417 513
FT DOMAIN: 516 606
FT DOMAIN: 609 708
FT DOMAIN: 711 822
FT DOMAIN: 825 916
FT DOMAIN: 918 1017
FT DOMAIN: 1030 1137
FT DOMAIN: 1147 1618
FT DOMAIN: 1619 1912
FT ACT_SITE: 1553 1553
FT ACT_SITE: 1844 1844
FT SITE: 1175 1178
FT CAEGHYE: 254 254
FT CARHGYD: 299 299
FT CARHGYD: 724 724
FT CARHGYD: 832 832
FT VARSPLIC: 181 189
FT VARSPLIC: 226 229
FT VARSPLIC: 775 783
FT VARSPLIC: 609 1137
FT MUTAGEN: 1178 1178
SC SIGNIFCTP 1913 AA, 214759 MW, 348090pI[182CF26 CRC64;

Query Match 11 58; Score 149. DB 1; Length 1912;
Best Local Similarity 21 58; Pred No 0 00032;
Matches 58; Conservative 37; Mismatches 90; Indels 62; Gaps 11;

QY 1 MCGPELLIVLVL-----PREMTSSVITVCKTFTVITIDTTPGSGASLCAVQNHTR 53
UB 2 VHAARLLLLILFLRLIQAHLPRKTRIVLGL-----VSGVASTICGATDPR 52
QY 54 EEELMTPEEYVPLDLSNNKINSSVAVSSISFNMLSTFPT-----GPIQS----- 102
UB 55 -PRIVN-----KKKKVSNQRT---EVLFEHDSGSVLHLPDTPRFAIYCVVA 100
QY 103 -----VSSVYLVNT-----EPPLLSGDKVQTEEGSNKLVNKKANPOQVMKY 148
UB 101 SNNVGLISVSKFVLVLEKEDLPKRFILIMSGPQKLVKRIKAIKLVAAASNPDELIMF 160
QY 149 KNSLLEDSKSHQVQVLTSESF-----VLSLKVKYKPNQVYSVLAASSLNLSLQ 199
UB 161 KQFLVDSNNKRLKSLNLSLSGIFIKALVLESELSGKATEVALNSAALVSAF 220
QY 200 FHLVND 206
DB 221 ANLYVRE 227

RESULT 13
LAMP_HUMAN STANDARD: PRT: 338 AA.
AC Q13449;
DT 01-NOV-1997 (rel. 35, last sequence update)
DT 01-NOV-1997 (rel. 35, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE Limbic system-associated membrane protein precursor (LSAMP).
GN LSAMP OR LAMP.
OS Homo sapiens (Human).

```



CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1]  
 CC RX MEDLINE=96235133; PubMed=8666243;  
 CC Pimenta A.F., Fischer I., Levitt P.;  
 CC \*cDNA cloning and structural analysis of the human limbic-system-  
 CC associated membrane protein (LAMP). \*;  
 CC RL Gene 170:189-195(1996).  
 CC -! FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.  
 CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF  
 CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).  
 CC -! SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC -! TISSUE SPECIFICITY: EXPRESSED ON LIMBIC NEURONS AND FIBER TRACTS  
 CC AS WELL AS IN SINGLE LAYERS OF THE SUPERIOR COLICULUS, SPINAL  
 CC CHORD AND CEREBELLUM.  
 CC -! SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY ISLN  
 CC SUBFAMILY.  
 CC -! SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: 041901; AAC50569.1; .  
 CC DR MIM: 603241; .  
 CC DR InterPro: IPR003006; Iq\_MHC.  
 CC DR InterPro: IPR003598; Iq\_C2.  
 CC DR InterPro: IPR003600; Iq\_Like.  
 CC DR Pfam: PF00047; Iq\_3.  
 CC DR SMART: SM00410; Iq\_Like\_1.  
 CC DR SMART: SM00408; IqC2\_2.  
 CC DR Immunoglobulin domain, Cell adhesion, Glycoprotein, GPI anchor,  
 CC Repeat, Signal.  
 CC KM Repeat: Signal.  
 CC FT SIGNAL 1 28 POTENTIAL.  
 CC FT CHAIN 29 315 LIMBIC SYSTEM-ASSOCIATED MEMBRANE  
 CC FT PROPEP 316 338 PROTEIN.  
 CC FT DOMAIN 46 118 IG-LIKE C2-TYPE DOMAIN 1.  
 CC FT DOMAIN 146 204 IG-LIKE C2-TYPE DOMAIN 2.  
 CC FT DISULFID 53 111 IG-LIKE C2-TYPE DOMAIN 3.  
 CC FT DISULFID 153 197 POTENTIAL.  
 CC FT DISULFID 239 290 POTENTIAL.  
 CC FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC FT LIPID 315 315 GPI-ANCHOR (POTENTIAL).  
 CC SQ SEQUENCE 338 AA: 37308 MM: 03455F286DF5D92F CRC64;  
 CC  
 CC Query Match 11.5%; Score 148.5; DB 1; Length 338;  
 CC Best Local Similarity 25.4%; Pred. No. 3; Be-05;  
 CC Matches 60; Conservative 34; Mismatches 97; Indels 45; Gaps 12;  
 CC  
 CC QY 2 QMKRLLVILFLPREMTSSVLTNGKENTILDTTSCASLCAVONHTREELLWYR 61  
 CC 11  
 CC DB 11 QAPLVLRLLCLPTGLRSVDFNRGDN--ITVRQDDTAIRLCVLED--KNSKVAMLN 66  
 CC 11  
 CC QY 62 EGS-----KVLKNSGKLNSSVCSVSSISLNQISFTCLHLSMVSVS-V 107  
 CC 11  
 CC DB 67 RSGIIFAGHDKMSLDPRVLELRHSI--EYSLRQKVNVTGDS--SYTSVQTQHEPKTSQV 124  
 CC 11

CC QY 108 VLVNTFPPELLSG-NDEQIVIEGSSVKLVNKANVQAMMYKNSSLDLEKSRHOIQQT 166  
 CC 11  
 CC DB 125 YLIVQVPPKISNISSDYVNEGSSVNTLVCMANGCPREPIITM-----RH-LTPT 171  
 CC 11  
 CC QY 167 SSSFC---LSITKVEKPEKGTSTAKSSLKTESLPHLYKENTVSPPIET 217  
 CC 11  
 CC DB 172 GREFGEVEVELELIDITKEUSGKEYC--KAANEVSAD---VKQKVVTNVPPT 221  
 CC 11  
 CC  
 CC RESULT 14  
 CC PGM: HUMAN  
 CC TP PGM: HUMAN STANDARD: PRT: 4393 AA.  
 CC AC P98160; Q16287;  
 CC PT 61-101-1346 (rel 14, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC UT 01-MAR-2002 (Rel. 41, Last annotation update)  
 CC DE Basament membrane-specific heparan sulfate proteoglycan core  
 CC DE protein precursor (HSPG) (Perlecan) (PLC).  
 CC GN HSPG2.  
 CC OS Homo sapiens (Human).  
 CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1]  
 CC RX MEDLINE=92112994; PubMed=1730768;  
 CC KA Kallunki P., Tryggvason K.;  
 CC \*Human basament membrane heparan sulfate proteoglycan core protein: a  
 CC 467-kD protein containing multiple domains resembling elements of the  
 CC low density lipoprotein receptor, laminin, neural cell adhesion  
 CC molecules, and epidermal growth factor.\*;  
 CC RL J. Cell Biol. 116:559-571(1992).  
 CC RM [2]  
 CC SEQUENCE FROM N.A.  
 CC RP TISSUE=Skin, and Colon;  
 CC RC MEDLINE=92235084; PubMed=1569102;  
 CC RX Murofuch A.D., Dodge G.R., Cohen J., Tuan P.S., Iozzo R.V.;  
 CC PA \*Primary structure of the human heparan sulfate proteoglycan from  
 CC basement membrane (HSPG2/perlecan). A chimeric molecule with multiple  
 CC domains homologous to the low density lipoprotein receptor, laminin,  
 CC RT RT J. Biol. Chem. 267:8544-8557(1992).  
 CC RN [3]  
 CC SEQUENCE OF 1918-1472 FROM N.A.  
 CC RP TISSUE=Colon;  
 CC RC MEDLINE=91365376; PubMed=1679749;  
 CC RX Dodge G.R., Kovalsky I., Chu M.L., Hassell J.R., McBride O.W.,  
 CC RA Yl H.F., Iozzo R.V.;  
 CC PA \*Heparan sulfate proteoglycan of human colon: partial molecular  
 CC cloning, cellular expression, and mapping of the gene (HSPG2) to the  
 CC short arm of human chromosome 1.\*;  
 CC RT RT Genomics 10:673-680(1991).  
 CC RN [4]  
 CC SEQUENCE OF 892-1398 FROM N.A.  
 CC RP TISSUE=Endothelium;  
 CC RC MEDLINE=92120660; PubMed=1685141;  
 CC RX Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,  
 CC RA Tryggvason K.;  
 CC PA \*Cloning of human heparan sulfate proteoglycan core protein,  
 CC assignment of the gene (HSPG2) to 1p36.1--p35 and identification of  
 CC a BamHI restriction fragment length polymorphism.\*;  
 CC RT RT Genomics 11:389-396(1991).  
 CC RN [5]  
 CC SEQUENCE OF 1-21 FROM N.A.  
 CC RP MEDLINE=94052171; PubMed=8234307;  
 CC RX Cohen I.R., Gaessel S., Murdoch A.D., Iozzo R.V.;  
 CC PA \*Structural characterization of the complete human perlecan gene and  
 CC its promoter.\*;  
 CC RT RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).  
 CC CC -! FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT  
 CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC  
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE MITRAPHILTRATION



Db 2542 SSSASLANGHT----LDLN-----CLVASQAPHT----ITWYKRGG--SLPSRHQIVGSR 2586  
 QY 79 VCVSSISENDGISFTCL-----GRDQSVSV-----VNTVFPPLLSNDQOTVE 126  
 Db 2587 LRIPQVTPADSG-PEYCHVNSGAGSRETSLIVTIQSGSSHYRVRV-PPRIREFSSPTIV 2644  
 QY 127 EGSNVKLVGNKANKPQOMMYKNSLLDEKSRHQIQGTSESFSQLSITKVEKPDNGTVS 186  
 Db 2645 ECGRLDNCVYARQPAIITWYKRGSL---PSRHQ---THGSHRLHQMSVADSGEYV 2697  
 QY 187 CLAKSLKTESLDFHLIVK-DKTYGVPIEP 215  
 Db 2698 CRANNNI--DALEASIVISPSKAGSPSAP 2725  
 RESULT 15  
 ID DCC\_MOUSE STANDARD; PRT: 1447 AA.  
 AC P70211;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Tumor suppressor protein DCC precursor.  
 DCC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-Brain;  
 RX MEDLINE=96112625; PubMed=8570174;  
 RA Cooper H.M., Arnes P., Britto J., Gad J., Wilks A.F.;  
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer  
 gene (DCC) and its expression in the developing mouse embryo.";  
 RL Oncogene 11:2243-2254(1995).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN-BALB/C; TISSUE-Brain;  
 RA Cooper H.M.;  
 RL Submitted (JUN-1996) to the FMBI/GenBank/EMBL databases.  
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM  
 CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A FLUPE  
 CC FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY  
 CC ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE  
 CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN  
 CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.  
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS  
 CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION  
 CC AND REMAIN AT THIS LEVEL IN THE ADULT.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X85788; CA59786.1; -  
 CC HSSP: P56276; ITLK.  
 DR MGD: MGI:94869; DCC.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003962; FNIII\_repeat.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003596; Ig\_c2.

DR InterPro: IPR003600; Ig\_Like.  
 DR Pfam: PF00041; fn3; 6.  
 DR Pfam: PF00047; Ig; 4.  
 DR PRINTS: PR00014; FNTPFIIII.  
 DR SMART: SM00060; FN3; 6.  
 DR SMART: SM00410; Ig\_Like; 2.  
 DR SMART: SM00408; IGC2; 3.  
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;  
 KW Repeat, Anti-oncogene; Alternative initiation; Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1447  
 FT  
 FT CHAIN 85 1447  
 FT  
 FT CHAIN 85 85  
 FT INIT\_MET 26 1097  
 FT DOMAIN 1098 1122  
 FT TRANSMEM 1123 1447  
 FT DOMAIN 34 124  
 FT DOMAIN 154 219  
 FT DOMAIN 254 317  
 FT DOMAIN 345 407  
 FT DOMAIN 426 522  
 FT DOMAIN 525 618  
 FT DOMAIN 619 716  
 FT DOMAIN 722 816  
 FT DOMAIN 840 940  
 FT DOMAIN 941 1042  
 FT DISULFID 61 117  
 FT DISULFID 161 212  
 FT DISULFID 261 310  
 FT DISULFID 352 400  
 FT CARBOHYD 60 60  
 FT CARBOHYD 94 94  
 FT CARBOHYD 299 299  
 FT CARBOHYD 318 318  
 FT CARBOHYD 478 478  
 FT CARBOHYD 628 628  
 FT CARBOHYD 702 702  
 FT VARSPIC 819 838  
 FT SEQUENCE 1447 AA; 158298 MW; 0D1F097C22D5B9F CRC64;  
 Query Match 11.4%; Score 147; DB 1; Length 1447;  
 Best Local Similarity 25.8%; pred. No. 0.00032;  
 Matches 51; Conservative 37; Mismatches 86; Indels 24; Gaps 9;  
 QY 11 ILPLPKEMISSVLTVNKNTENY;LDTTPSSQASLCAVNHRTREELLYREEGRYDLKS 70  
 Db 240 LYFLQRR--PSNVIAIEEK-----DAVLECCVSGYP-PPSFTMLRGHEVYQLRS 284  
 QY 71 G--MKINSSVCVSSISENDGISFTCLG--RDQSVSVVAVLWVTPP-LISGNDQOTVE 126  
 Db 285 KRYSLGSGNLLISNVITDDDSG--TYTCVTVYKKNENISASBELVLPVPEFLNPSNLVAY 343  
 QY 127 EGSNVKLVGNKANKPQOMMYKNSLLDEKSRHQIQGTSESFSQLSITKVEKPDNGTVS 186  
 Db 344 ESMIDIEFCVAGSKPVPVPMNMKNKGDDV-IPSDYFOIVGSS---NLRILGVVKSDEGFGY 399  
 QY 187 CLAKSLKTESLDFHLIV 404  
 Db 400 CVANEAGCAQSSAOLIV 417

Search completed: September 17, 2002, 13:24:15  
 Job time: 802 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:24:15 : Search time 40.16 Seconds

(without alignments)  
252.603 Million cell updates/sec

Title: US-09-863-823-6

Sequence: 1336  
1 MAWKSSVIMQMGKRFLLVIL.....RRKIMKLCMKDKPHESTAL 262

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	176	13.2	853	1 NCAL_BOVIN	P18186 Bos taurus
2	175	13.1	858	1 NCAL_PAT	P13596 rat'us norv
3	173	12.9	359	1 LACH_DROME	O24372 drosophila
4	170	12.7	2912	1 DSCA_HUMAN	O60469 homo sapien
5	166	12.4	761	1 NCAL_HUMAN	P13599 homo sapien
6	166	12.4	848	1 NCAL_HUMAN	P13599 homo sapien
7	160.5	12.0	1091	1 NCAL_CHICK	P13590 gallus gall
8	159	11.9	725	1 NCAL_MOUSE	P13594 mus musculu
9	159	11.9	1115	1 NCAL_MOUSE	P13594 mus musculu
10	153	11.5	2029	1 LAR_DROME	P16621 drosophila
11	150	11.2	1912	1 PTPD_HUMAN	P23468 homo sapien
12	149	11.2	1447	1 DCC_HUMAN	P43146 homo sapien
13	148.5	11.1	338	1 LAMP_HUMAN	O13449 homo sapien
14	147.5	11.0	4393	1 PGRM_HUMAN	P29166 homo sapien
15	147	11.0	1447	1 LOC_MOUSE	P70211 mus musculu
16	145	10.9	837	1 NCAL_MOUSE	O15394 mus musculu
17	144.5	10.8	338	1 LAMP_CHICK	O98919 gallus gall
18	144.5	10.8	338	1 LAMP_RAT	O62817 rattus norv
19	144.5	10.8	1348	1 VGR2_CHICK	P32584 rattus norv
20	144	10.8	521	1 CEAL_MOUSE	P31809 mus musculu
21	142.5	10.7	345	1 LACH_SCHAM	U60474 schistocor
22	142	10.6	837	1 NCAL_MOUSE	O15176 mus musculu
23	142	10.6	1356	1 VGR2_HUMAN	P35968 homo sapien
24	141.5	10.6	1493	1 NCAL_MOUSE	O02173 mus musculu
25	139	10.4	1450	1 MRSF_CHICK	P02173 mus musculu
26	137	10.3	1092	1 NCAL_CHICK	P36333 xenopus lae
27	137	10.3	1343	1 VGR2_RAT	O08173 rattus norv
28	136.5	10.2	519	1 ECTO_RAT	P16573 rattus norv
29	136.5	10.2	1897	1 PPRF_HUMAN	P10586 homo sapien
30	135.5	10.1	1070	1 PTK7_HUMAN	U13308 homo sapien
31	135	10.1	1377	1 NCAL_RAT	P97603 rattus norv
32	134	10.0	344	1 CEAL_HUMAN	P40194 homo sapien
33	133.5	10.0	3707	1 PGBM_MOUSE	U05793 mus musculu

34	132.5	9.9	976	1 KIT_HUMAN	P10721 homo sapien
35	131.5	9.8	353	1 CEPD_CHICK	O90773 gallus gall
36	131.5	9.8	526	1 NCAL_HUMAN	P13688 homo sapien
37	131	9.8	1044	1 NCAL_XENLA	P16170 xenopus lae
38	130.5	9.8	298	1 JAM2_HUMAN	P57067 homo sapien
39	130.5	9.8	702	1 CEAL_HUMAN	P06731 homo sapien
40	130.5	9.8	864	1 TRKC_RAT	O03351 rattus norv
41	130	9.7	1367	1 VGR2_MOUSE	P35918 mus musculu
42	129.5	9.7	333	1 AMAL_DROME	P15364 drosophila
43	129.5	9.7	626	1 MAJ_MOUSE	P20917 mus musculu
44	129.5	9.7	822	1 FCRL_HUMAN	P11362 homo sapien
45	129.5	9.7	822	1 FCRL_MOUSE	P14052 mus musculu

## ALIGNMENTS

RESULT 1  
ID NCAL\_BOVIN STANDARD: PRT: 853 AA.  
AC P18186;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
WT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Neural cell adhesion molecule, 140 kDa isoform precursor (N-CAM 140)  
DE (NCAM-140).  
GN NCAM1 OR NCAM.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OC NCBI\_Taxid=9913;  
RN 1  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Brain cortex;  
RX MEDLINE=89378239; PubMed-2776887;  
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshayakov M.V.,  
RA Petukhova G.V., Rakitina T.V., Feshchenko E.A., Ishchenko K.A.,  
RA Mirzoeva S.F., Chernova M.N., Dranytsyna S.M.,  
RT Calmodulin-independent bovine brain adenylate cyclase. Amino acid  
RT sequence and nucleotide sequence of the corresponding cDNA.\*;  
PI FEBS Lett 254:69-74(1989).  
RN 2  
RP SEQUENCE OF 20-36.  
RX MEDLINE=86140120; PubMed-3512556;  
RA Rougon G., Marshaq D.R.,  
RT Structural and immunological characterization of the amino-terminal  
RT domain of mammalian neural cell adhesion molecules.\*;  
PI J Biol Chem 261:3396-3401(1986).  
RN 3  
RP IDENTIFICATION AS N-CAM.  
RX MEDLINE=92111748; PubMed-1765159;  
RA Fremont R.T.,  
RT "A bovine brain cDNA purifying to enzyme calmodulin-insensitive  
RT adenylate cyclase has extensive identity with neural cell adhesion  
RT molecules (N-CAMs)."  
PI FEBS Lett 295:230-231(1991).  
RN 4  
RP FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
RN NEURON-NEURON ADHESION, NEURITE PROLIFERATION, OUTGROWTH OF  
RN NEURITES, ETC.  
CC 1. SUBCELLULAR LOCATION: Type 1 membrane protein.  
CC 1. ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF  
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC 1. SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC 1. SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC 1. SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
CC 1. CAUTION: WAS ORIGINALLY (REF. 1) THOUGHT TO BE A CALMODULIN-  
CC INDEPENDENT ADENYLATE CYCLASE.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way



FT CARBOHYD 316 316 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 348 348 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 434 434 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 460 460 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 489 489 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 858 AA: 94658 MW: EALAD6M4EAD50F6 CRC64;

Query Match 13.1%; Score 175; DB 1; Length 858;  
 Best Local Similarity 28.4%; Pred. No. 1 1e-06;

Matches 60; Conservative 31; Mismatches 88; Indels 32; Gaps 8;

QY 49 QASLIC-----AVQNRRELLMYRECPVDLKSQKINS---SSVYS 90  
 DB 325 QVLTCEASGPISITWRTSTRNISSEKASWTRPEKQETLDGHVYRSHARVSSLTK 384  
 QY 91 SISENDNGISFTC---RLGRQSVSVVLYNTFPLLSGNDFOVEGSKVLCVNYK 146  
 DB 385 SIQYTDAG-EYICIASNIGD---SOSMYLEVQARKLQGPAAVYTWEGNQVNTICEV 440  
 QY 147 ANPOAQMWMYNNSSLLDLK-SRHQIQOTESFOLSTIKVEKPDNGYSCIAKSLKTES 205  
 DB 441 AYPASATISMFRDGLSSNSYNIKITYNPSASYLEVTPDSENDFGYNGCTAVNRIGQES 500  
 QY 206 LDFHLIVKDKTVGP---LEPLIACVVF 232  
 DB 501 LEFLLVQAD-TPSSPSIDRVPYSSTAQVGF 510

## RESULT 3

LACH\_DROME STANDARD; PRT: 359 AA.

ID LACH\_DROME  
 AC Q24372.  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Lachesis precursor.  
 GN LAC.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94038693; PubMed=8223276;  
 RA Karlstrom R.O., Wilder L.P., Bastiani M.J.;  
 RT "Lachesis: an immunoglobulin superfamily protein whose expression  
 RT correlates with neurogenesis in grasshopper embryos.";  
 RL Development 118:509-522(1993).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND  
 CC AXON OUTGROWTH.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By  
 CC similarity).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN DIFFERENTIATING NEURONAL CELLS  
 CC FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL  
 CC NERVOUS SYSTEMS.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY  
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L13355; AAC37184.1; -  
 DR HSSP: P56276; ITLK.  
 DR Flybase: FBgn0010238; LAC.  
 DR Interpro: IPR003006; Ig\_MHC.

DR Interpro: IPR003598; Ig\_C2.  
 DR Interpro: IPR003600; Ig\_Like.  
 DR Pfam: PF00047; Ig\_3.  
 DR SMART: SM00410; Ig\_Like; 1.  
 DR SMART: SM00408; IC52; 2.  
 KM Immunoglobulin domain, Cell adhesion, Glycoprotein, GPI anchor;  
 Repeat: Signal;  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 336 LACHESIN.  
 FT PROPEP 337 354 FEMVELL IN MAJORE FORM (POTENTIAL).  
 FT DOMAIN 43 120 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 150 211 IG-LIKE C2-TYPE DOMAIN 1.  
 FT DOMAIN 240 310 IG-LIKE C2-TYPE DOMAIN 2.  
 FT DISULFID 50 113 POTENTIAL.  
 FT DISULFID 157 204 POTENTIAL.  
 FT DISULFID 247 303 POTENTIAL.  
 FT CARBOHYD 92 92 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT LIPID 336 336 GPI-ANCHOR (POTENTIAL).  
 SQ SEQUENCE 359 AA: 39998 MW: DIF513E2B9D86E6 CRC64;

Query Match 12.9%; Score 173; DB 1; Length 359;  
 Best Local Similarity 27.3%; Pred. No. 5.2e-07;

Matches 60; Conservative 43; Mismatches 85; Indels 32; Gaps 11;

QY 1 WSSVIMQMFFLLVILFLPEMTSSVLVW-NFENVLLTTP-NASLI-AYQNHTE ~2  
 DB 2 WPSISNVM-LELIAL-FVQNLACPLISVIFKIMID--SVFELASVQ-YAKE 57  
 QY 63 EELLMYR-FEGPVLKSNKI-----NSS--VCYSSISFNNGISFTCR-- 104  
 DB 58 YNVEFLKTDSPFPLSTSTVLVIKDSRFLKYPDNSSYTLQIKDIQETDAG-TYTCQVV 116  
 QY 105 LGRDQSVSVVLYNTFPLLSGNDFOV--EEGSKVLCVNYKANPOAQMWMYK-NSSL 161  
 DB 117 ISTVKAVALKLSVKRKPVALSDNSISVVAESBQVMGCVASGYPTPTITRRRNNAI 176  
 QY 162 LLEKSRHQIQOTESFOLSTIKVEKPDNGYSCIAKSSL 201  
 DB 177 LPTDSA-----TYVGNLTPIKSVKKEPDGYVCVADNGV 210  
 RESULT 4  
 DSCA\_HUMAN STANDARD; PRT: 2012 AA.  
 ID DSCA\_HUMAN  
 AC Q06489; Q06489;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Down syndrome cell adhesion molecule precursor (CHD2).  
 GN DSCAM.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RC TISSUE-Brain.  
 RX MEDLINE=98087574; PubMed=9426258;  
 RA Yanakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,  
 RA Lyons G.E., Korenberg J.R.;  
 RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a  
 RT down syndrome region and is involved in the development of the  
 RT nervous system.";  
 RL Hum. Mol. Genet. 7:227-237(1998).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20289799; PubMed=10830953;  
 RA Hattori M., Toyiyama A., Taylor T.D., Watanabe H., Yada T.,  
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
 RA Menzel U., Delabau J., Kumpf K., Lehmann K., Patterson D.,











RESULT 8  
NCAM\_MOUSE  
ID NCAM\_MOUSE STANDARD: PKT: 725 AA.  
AC P13595; Q61950;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Neural cell adhesion molecule, Phosphatidylinositol-linked isoform  
DE Precursor (N-CAM 120) (NCAM-120).  
GN NCAM1 OR NCAM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Scleroptalia; Muridae; Mus.  
OX NCBI\_TaxID:10090;  
RN [1]  
RP SOURCE FROM N.A.  
RC STRAIN:G57BL/6;  
KA MEDLINE:87246524; PubMed:3595563;  
KA Bittels D., Santoni M.-L., Willie W., Ruppert G., Calix J.-C.,  
KA Hirsch M.-R., Fontecilla-Camps J.-C., Coridis G.,  
KT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for  
a Mr 79,000 polypeptide without a membrane-spanning region."  
KT EMBO J. 6:907-914(1987).  
RL [2]  
RN SOURCE: 23 750 FROM N.A.  
RC STRAIN:G57BL/6; TISSUE:Brain;  
KA MEDLINE:8923563; PubMed:2721486;  
KA Santoni M., Bittels D., Ruppert G., Horod A., Coridis G., Willie W.,  
KT "Differential exon usage involving an unusual splicing mechanism  
generates at least eight types of NCAM cDNA in mouse brain."  
KT EMBO J. 8:385-392(1989).  
RL [1]  
RN SEQUENCE OF 642-725 FROM N.A.  
KA MEDLINE:88283628; PubMed:3396534;  
KA Barbas J.A., Chaux J.C., Steinmetz M., Gattis G.,  
KT "Differential splicing and alternative polyadenylation generates  
distinct NCAM transcripts and proteins in the mouse."  
KT EMBO J. 7:625-632(1988).  
RL [4]  
RN SEQUENCE OF 20-46.  
KA MEDLINE:86140170; PubMed:4512556;  
KA Koupou G., Marshak D.R.,  
KT "Structural and immunological characterization of the amino-terminal  
domain of mammalian neural cell adhesion molecules."  
KT J. Biol. Chem. 261:3396-3401(1986).  
CC - FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
NEURITES, ETC.  
CC - SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC - ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF  
N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC - SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
CC - SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
CC - SIMILARITY: CONTAINS 2 FIBRONECTIN-TYPE III-LIKE DOMAINS.  
CC  
CC This SwissProt entry is a highlight. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb.slb.ch/announce/>  
or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
CC  
EMBL: Y00051; CAA68261.1;  
EMBL: X15049; CAA33148.1; ALT\_SEQ.  
EMBL: X07195; CAA30174.1;  
PIR: A29674; TMSN;  
MOL: M61-97281; Neam.  
InterPro: IPR003961; FN\_111.  
InterPro: IPR003006; Iq\_MIG.  
InterPro: IPR003598; Iq\_C2.  
Pfam: PF00041; Iq3; 2.

PIfam: PF00047; Iq; 5.  
DR SMART: SM00060; FN3; 2.  
DR SMART: SM00408; ICG2; 5.  
KW Cell adhesion, glycoprotein, heparin, Alternative splicing;  
KW Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.  
FT STGNAL 1 19  
FT CHAIN 20 725  
FT  
FT DOMAIN 34 103  
FT DOMAIN 132 196  
FT DOMAIN 228 293  
FT DOMAIN 323 393  
FT DOMAIN 420 487  
FT DOMAIN 519 596  
FT DOMAIN 625 692  
FT DOMAIN 152 156  
FT DOMAIN 161 165  
FT DISULFID 41 96  
FT DISULFID 139 189  
FT DISULFID 235 288  
FT DISULFID 330 386  
FT DISULFID 427 480  
FT CARBOHYD 222 222  
FT CARBOHYD 316 316  
FT CARBOHYD 348 348  
FT CARBOHYD 424 424  
FT CARBOHYD 450 450  
FT CARBOHYD 479 479  
FT CONFLICT 263 368  
FT CONFLICT 273 273  
FT CONFLICT 354 355  
FT CONFLICT 549 549  
FT CONFLICT 572 572  
FT CONFLICT 575 575  
FT CONFLICT 589 594  
FT CONFLICT 602 602  
FT CONFLICT 657 657  
SO SEQUENCE 725 AA; 80296 MW; C2A8R8H4461C6B2P CMO64;  
Query Match 11.98; Score 159; DB 1; Length 725;  
Best Local Similarity 27.58; Pred. No. 1.8e-05;  
Matches 57; Conservative 35; Mismatches 81; Indels 34; Gaps 9;  
UY 49 CASLTC-----AVGNHREELHMYRKECHVLDKSNKINSSVSSLSK 94  
DB 325 LVITTCASGPIPTSTMTKSTINISSEFQHL---TCHMVVSHARY--SLTLKSLQY 378  
UY 95 NDNCSLFTC-----KLDKDSVSVSVLVNITFPLLSGNSNPEUVEGGSVNVKLVNKNAPV 150  
DB 379 IFAG-EYMTASNT150F---SSSHIFQYAPKIGFPAVVTMGVNVITFVPAVPS 434  
UY 151 AQMMWVNSNLSLIDPE-SPLDQGLSSSEYSLIVENKSNVYVPSKRTFSTFQ 209  
DB 445 ATTSWPEFGLTSSVSNSTFTNINPACVIVIPSENFQANVWLAIVNLDQFSLHFI 494  
UY 213 LVEIKKSVLE IEFIAACVVF 232  
DB 495 LVQAD-IFSSSIDRVEYSSTAQVGF 520  
RESULT 9  
NCAM\_MOUSE  
ID NCAM\_MOUSE STANDARD: PKT: 1115 AA.  
AC P13595; Q61949;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Neural cell adhesion molecule, large isoform precursor (N-CAM 180)  
DE (NCAM-180) [Contains: N-CAM 140 (NCAM-140)].  
GN NCAM1 OR NCAM.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

CC Mammalia, Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (N-CAM 180).  
 RC STRAIN-C57BL/6;  
 RX MEDLINE=87246524; PubMed=3595563;  
 RA Barthele D., Santoni M.J., Wille W., Rupert C., Chai J.C.,  
 Hirsch M.R., Fontecilla-Camps J.C., Goridis C.;  
 RT "Isolation and nucleotide sequence of mouse N-CAM cDNA that codes for  
 a Mr 79,000 polypeptide without a membrane-spanning region.";  
 RL EMBL J. 6:907-914(1987).  
 RN [2]  
 RP SEQUENCE OF 529-1115 FROM N.A. (N-CAM 140).  
 RC STRAIN-C57BL/6;  
 RX MEDLINE=88067687; PubMed=3684567;  
 RA Santoni M.-J., Barthele D., Barbas J.A., Hirsch M.-P., Steinmetz M.,  
 Goridis C., Wille W.;  
 RT "Analysis of cDNA clones that code for the transmembrane forms of the  
 mouse neural cell adhesion molecule (NCAM) and are generated by  
 RT alternative RNA splicing.";  
 RL Nucleic Acids Res. 15:8621-8641(1987)  
 RN [3]  
 RP SEQUENCE OF 642-1115 FROM N.A. (N-CAM 180)  
 RX MEDLINE=88283628; PubMed=3396534;  
 RA Barbas J.A., Chai J.C., Steinmetz M., Goridis C.;  
 RT "Differential splicing and alternative polyadenylation generates  
 RT distinct NCAM transcripts and proteins in the mouse.";  
 RL EMBL J. 7:625-632(1988).  
 RN [4]  
 RP SEQUENCE OF 804-1081 FROM N.A. (N-CAM 180)  
 RC STRAIN-C57BL/6J; Tissue=Brain;  
 RX MEDLINE=88247737; PubMed=2454455;  
 RA Barthele D., Vopper G., Wille W.;  
 RT "NCAM-180, the large isoform of the neural cell adhesion molecule of  
 RT the mouse, is encoded by an alternatively spliced transcript.";  
 RL Nucleic Acids Res. 16:4217-4225(1988)  
 RN [5]  
 RP SEQUENCE OF 702-1115 FROM N.A. (N-CAM 140)  
 RC STRAIN-C57BL/6; Tissue=Brain;  
 RX MEDLINE=89251563; PubMed=2721486;  
 RA Santoni M.J., Barthele D., Vopper G., Boned A., Goridis C., Wille M.;  
 RT "Differential exon usage involving an unusual splicing mechanism  
 RT generates at least eight types of NCAM cDNA in mouse brain.";  
 RL EMBL J. 8:385-392(1989).  
 RN [6]  
 RP SEQUENCE OF 20-36.  
 RX MEDLINE=86140120; PubMed=3512556;  
 RA Rougon G., Marshak D.R.;  
 RT "Structural and immunological characterization of the amino-terminal  
 RT domain of mammalian neural cell adhesion molecules.";  
 RL J Biol Chem 261:3396-3401(1986)  
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
 CC NEURITES, ETC.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF  
 CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X07200; CAA30177.1; -;  
 DR EMBL: Y00051; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: X06328; CAA29641.1; -;  
 DR EMBL: X07195; CAA30173.1; -;

DR EMBL: X07244; CAA30230.1; -;  
 DR EMBL: X15051; CAA33150.1; -;  
 DR EMBL: X15052; CAA33151.1; -;  
 DR PIR: A29673; IJMSNL.  
 DR MGI: 97281; Ncam.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR Pfam: PF00041; In3; 2.  
 DR Pfam: PF00047; Ig; 5.  
 DR SMART: SM0060; FN3; 2.  
 DR SMART: SM00408; IgC2; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;  
 KW Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.  
 FT CHAIN 1 19  
 FT 20 1115  
 FT DOMAIN 20 711  
 FT TRANSEM 712 729  
 FT DOMAIN 730 1115  
 FT DOMAIN 34 103  
 FT DOMAIN 132 196  
 FT DOMAIN 228 295  
 FT DOMAIN 323 393  
 FT DOMAIN 420 487  
 FT DOMAIN 519 596  
 FT DOMAIN 625 692  
 FT DOMAIN 152 165  
 FT DOMAIN 161 165  
 FT DISULFID 41 94  
 FT DISULFID 139 189  
 FT DISULFID 235 288  
 FT DISULFID 330 386  
 FT DISULFID 427 480  
 FT CARBOHYD 222 222  
 FT CARBOHYD 316 316  
 FT CARBOHYD 348 348  
 FT CARBOHYD 424 424  
 FT CARBOHYD 450 450  
 FT CARBOHYD 479 479  
 FT ASPGLIC 810 1076  
 SQ SEQUENCE 1115 AA; 119351 MW; 2C93DCD474CFBCAF CRC64;  
 Query Match 11.9%; Score 159; DB 1; Length 1115;  
 Best Local Similarity 27.5%; Pred. No. 3; Le-05;  
 Matches 57; Conservative 35; Mismatches 81; Indels 34; Gaps 9;  
 QY 49 QASLIG-----AVONHTREELMYREGRVVDLSGNKINSSVCVSSISE 94  
 DB 325 QVTLTCAASDPITPTSTPTSTNINSSREGDL----DGHMVPSHARV--SSLTLKSIQY 378  
 QY 95 NDNGISFTC---RLGDSVSYSVVALNTPPLLSGNDFQVEBSNKLVCNKNKANO 150  
 DB 379 RDAG-EYMCASANTIGD---SOSIDFQYAPKLOGPAVYTWECNVCNITCEFAVPS 434  
 QY 151 AGMMWYNYNSLELEK-SHU-LWISF-LSLIKVEKIKNGIYS-LASLSLLESLFPH 209  
 DB 435 ATISWFDQQLPSSNSNINIKITNPASYLEVTPQSEDFGNKCTANVPVIGDSLEPI 494  
 QY 210 LIVDKTVGVV---IEPIIACVVF 232  
 DB 495 LVQAD-TPSPSIDRVEPYSSTAQVQ 520  
 RESULT 10  
 LAR DRONE STANDARD; PRT; 2029 AA.  
 AC P16621;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase lar precursor (EC 3.1.3.48) (Protein-



CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 CC NCRI\_TaxID-9606;  
 CC [1]  
 CC SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.  
 CC MEDLINE-95204468; PubMed-7896816;  
 CC Pullido K., Krieger N.X., Seta-Pages C., Saito H., Streuli M.;  
 CC "Molecular characterization of the human transmembrane protein-  
 CC tyrosine phosphatase delta. Evidence for tissue-specific expression of  
 CC alternative human transmembrane protein-tyrosine phosphatase delta  
 CC isoforms.";  
 CC J. Biol. Chem. 270:6722-6728(1995).  
 CC [2]  
 CC SEQUENCE OF 390-1912 FROM N.A.  
 CC TISSUE-Placenta;  
 CC MEDLINE-91006018; PubMed-2170109;  
 CC Krueger N.X., Streuli M., Saito H.;  
 CC "Structural diversity and evolution of human receptor-like protein  
 CC tyrosine phosphatases.";  
 CC EMBO J. 9:3241-3252(1990).  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> Protein  
 CC tyrosine + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT  
 CC TISSUES DUE TO ALTERNATIVE SPLICING.  
 CC -1- PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN  
 CC FROM THE TRANSMEMBRANE SEGMENT.  
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: L38929; AAC4749.1;  
 CC EMBL: X54133; CAA8068.1;  
 CC PIR: S12052; S12052.  
 CC HSP: P18052; LYFO.  
 CC MIM: 601598;  
 CC InterPro: IPR003961; FN\_III.  
 CC InterPro: IPR003962; FNIII\_repeat.  
 CC InterPro: IPR003006; Ig\_MHC.  
 CC InterPro: IPR003598; Ig\_C2.  
 CC InterPro: IPR000387; TYR\_phosphatase.  
 CC InterPro: IPR000242; TYR\_prot\_phphatase.  
 CC Pfam: PF00047; Ig\_3.  
 CC Pfam: PF00102; Y\_phosphatase\_2.  
 CC PRINTS: PR00014; FNYPEI11.  
 CC PRINTS: PR00700; PRTYPHPTASE.  
 CC SMART: SM00060; FN3; 8  
 CC SMART: SM00408; IGC2; 3.  
 CC SMART: SM00194; PTPC; 2.  
 CC PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.  
 CC PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.  
 CC PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 CC K1M Hydroxylase Receptor; Glycoprotein; Signal; Transmembrane; Repeat;  
 CC Immunoglobulin domain; Alternative splicing.  
 CC FT SIGNAL 1 20 POTENTIAL.  
 CC FT CHAIN 1 1912 PROTEIN-TYROSINE PHOSPHATASE DELTA.  
 CC FT DOMAIN 21 1265 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 1266 1290 POTENTIAL.  
 CC FT DOMAIN 1291 1912 CYTOPLASMIC (POTENTIAL).  
 CC FT DOMAIN 23 115 IG-LIKE C2-TYPE DOMAIN 1.  
 CC FT DOMAIN 118 225 IG-LIKE C2-TYPE DOMAIN 2.  
 CC FT DOMAIN 232 318 IG-LIKE C2-TYPE DOMAIN 3.  
 CC FT DOMAIN 320 414 FIBRONECTIN TYPE-III 1  
 CC FT DOMAIN 417 513 FIBRONECTIN TYPE-III 2

FT DOMAIN 516 606 FIBRONECTIN TYPE-III 3.  
 FT DOMAIN 609 708 FIBRONECTIN TYPE-III 4.  
 FT DOMAIN 711 822 FIBRONECTIN TYPE-III 5.  
 FT DOMAIN 825 916 FIBRONECTIN TYPE-III 6.  
 FT DOMAIN 918 1017 FIBRONECTIN TYPE-III 7.  
 FT DOMAIN 1020 1137 FIBRONECTIN TYPE-III 8.  
 FT DOMAIN 1137 1618 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 1619 1912 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT ACT\_SITE 1553 1553 BY SIMILARITY.  
 FT ACT\_SITE 1844 1844 CLEAVAGE (POTENTIAL).  
 FT SITE 1175 1178 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 254 254 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 299 299 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 724 724 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 832 832 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSPLIC 181 189 MISSING (IN KIDNEY ISOFORM).  
 FT VARSPLIC 226 229 MISSING (IN KIDNEY ISOFORM).  
 FT VARSPLIC 775 783 MISSING (IN KIDNEY ISOFORM).  
 FT VARSPLIC 609 1137 MISSING (IN FETAL BRAIN ISOFORM).  
 FT MUTAGEN 1178 1178 R->A: 2.5-FOLD REDUCTION IN CLEAVAGE.  
 SQ SEQUENCE 1912 AA; 234759 MW; 3AEB0BCD3182E26 CRC64;  
 Query Match 11.2%; Score 150; DB 1; Length 1912;  
 Best Local Similarity 23.4%; Pred. No. 0.00033;  
 Matches 58; Conservative 38; Mismatches 90; Indels 62; Gaps 11;  
 QY 8 IMUGSEILLVILFL-----PREMTSSVLTVNCKTENYLLDTPGSQASTICAVONHT 60  
 LB 1 MHVAVLALVLELFLDLALVLPETIPVDLS-----VSQVASFICVAILGP 51  
 QY 61 YHFLILWYEESEVRLKSNINSSSVVSSISNWNLSITPL-----SDQS----- 110  
 DB 52 R-PEIYWN-----KKGKVSNOFP---EYIEFDSSGVLLPIQLPPEDEAIYECV 99  
 QY 111 -----VSQVYINVT-----FPLLSGNDFOYEESSNKLIVNKANVQAQMM 155  
 DB 100 ASNNWELISVSRILYVLRHEDQIPRGPIPIIDMSPOUKVVERKRTFMCAAGNIDPPIIW 159  
 QY 156 YNNSLLEKESPHQIQ-TSRSF-----GLSIKKYK\*KN\*SY\*TA\*~L\*LF~\* 206  
 DB 160 FKDFLPVDTSNNGNRKQKRSISIGTPIKCALQIEQSESDQKYECVATNSAGTYSA 219  
 QY 207 DPHLYVKD 214  
 DB 220 PAVLYVRE 227  
 RESULT 12  
 DCC\_HUMAN STANDARD; PRT; 1447 AA.  
 AC P43146;  
 DT 01-NOV-1995 (Rel. 32; Created)  
 DT 01-NOV-1995 (Rel. 32; Last sequence update)  
 DT 01-MAR-2002 (Rel. 41; Last annotation update)  
 DE Tumor suppressor protein p53 precursor (Colorectal cancer suppressor).  
 DE DCC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NCBI\_TaxID-9606;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE-95015323; PubMed-7826723;  
 CC Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,  
 CC Vogelstein B.;  
 CC "The DCC gene product in cellular differentiation and colorectal  
 CC tumorigenesis.";  
 CC Genes Dev. 8:1174-1183(1994).  
 CC [2]  
 CC SEQUENCE OF 1-750 FROM N.A.  
 CC MEDLINE-95015323; PubMed-7826723;  
 CC Fearon F.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,

[illegible]



CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1]  
 CC RP SEQUENCE FROM N.A.  
 CC MEDLINE=96235133; PubMed=8666243;  
 CC Pimenta A.F., Fischer I., Levitt P.;  
 CC "cDNA cloning and structural analysis of the human limbic system  
 CC associated membrane protein (LAMP).";  
 CC RT associated membrane protein (LAMP).";  
 CC RL Gene 170:189-195(1996).  
 CC -1- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.  
 CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF  
 CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH  
 CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).  
 CC -1- SUPPLEMENTARY LOCATION: Attached to the membrane by a GPI-anchor  
 CC -1- TISSUE SPECIFICITY: EXPRESSED ON LIMBIC NEURONS AND FIBER TRACTS  
 CC AS WELL AS IN SINGLE LAYERS OF THE SUPERIOR COLICULUS, SPINAL  
 CC CHORD AND CEREBELLUM.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>  
 CC or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).  
 CC -----  
 CC DR EMBL: U41901; AAC50569.1; -;  
 CC DR MIM: 603241; -;  
 CC DR InterPro: IPR003036; Iq\_MHC.  
 CC DR InterPro: IPR003598; Iq\_C2.  
 CC DR InterPro: IPR003600; Iq\_Like.  
 CC DR Pfam: PF00047; Iq\_3.  
 CC DR SMART: SM00410; Iq\_Like; 1.  
 CC DR SMART: SM00408; Iq\_C2; 2.  
 CC KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;  
 CC Repeat; Signal.  
 CC FT SIGNAL 1 28  
 CC FT CHAIN 29 315  
 CC FT PROPEP 316 338  
 CC FT DOMAIN 46 118  
 CC FT DOMAIN 146 204  
 CC FT DISULFID 232 297  
 CC FT DISULFID 53 111  
 CC FT DISULFID 153 197  
 CC FT DISULFID 239 290  
 CC FT CARBOHYD 40 40  
 CC FT CARBOHYD 66 66  
 CC FT CARBOHYD 136 136  
 CC FT CARBOHYD 148 148  
 CC FT CARBOHYD 279 279  
 CC FT CARBOHYD 287 287  
 CC FT CARBOHYD 300 300  
 CC FT CARBOHYD 315 315  
 CC FT LIPID 315 315  
 CC FT SEQUENCE 338 AA; 37308 MW; 03455F286DF592F CRC64;  
 CC  
 CC Query Match 11.1%; Score 148.5; DB 1; Length 338;  
 CC Best Local Similarity 25.4%; Pred. No. 4.8e-05;  
 CC Matches 60; Conservative 34; Mismatches 97; Indels 45; Gaps 12;  
 CC  
 CC 10 OMGRLLVILFLPREMTSSVLTNGKENTILDTTGSASALICAVONTREELLWR 69  
 CC 11 QLPVLRLCLDPLGPRVSNRPNRGTNN-TVRGQPTALLPVLLED--KNSKVAWLN 66  
 CC DB 70 EEG-----RVDKSKGNKINSSWYSSISSENGISSEFTRRLPDSQVSSV-V 115  
 CC DB 67 RSGIIFACHDKMSLDPRVELFKRHSIL-EYSILRIQKAVVTEDEG-SYTSQSVLTQMEKPTISLY 124

CC 116 VLANTFPLLSG-NDPFTVEBSNVKLVNKNFVAQMMKKNSSLDLEKSKRLQVOT 174  
 CC 125 YLIVQPPKISINSSDVTVEGNSVTLVCMANGRPPEVITW-----RH-LPT 171  
 CC 175 SESFC-----LSTKVEKPDNGTGYSCAKSSEKTESLDPHLVYKOKTVCPLEPI 225  
 CC 172 GRECEEEVEILIGITREOSCKTEC KANVESSAD-----VKQKVTYVNPPTI 221  
 CC  
 CC RESULT 14  
 CC PGHM\_HUMAN STANDARD; PRT; 4393 AA.  
 CC AC P98160; O16287;  
 CC DT 01-OCT-1996 (Pel. 34, Created)  
 CC DT 01-OCT-1996 (Pel. 34, Last sequence update)  
 CC DT 01-MAR-2002 (Pel. 41, Last annotation update)  
 CC DE Basal membrane-specific heparan sulfate proteoglycan core  
 CC protein precursor (HSPG2) (Perlecan) (PLC).  
 CC GN Homo sapiens (Human).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1]  
 CC RP SEQUENCE FROM N.A.  
 CC MEDLINE=92112944; PubMed=1730768;  
 CC Kallunki P., Trygvaason K.;  
 CC "Human basement membrane heparan sulfate proteoglycan core protein: a  
 CC 467-kD protein containing multiple domains resembling elements of the  
 CC low density lipoprotein receptor, laminin, neural cell adhesion  
 CC molecules, and epidermal growth factor.";  
 CC RT J. Cell Biol. 116:559-571(1992).  
 CC [2]  
 CC RP SEQUENCE FROM N.A.  
 CC TISSUE=Skin, and Colon;  
 CC MEDLINE=92235084; PubMed=1569102;  
 CC Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;  
 CC "Primary structure of the human heparan sulfate proteoglycan from  
 CC basement membrane (HSPG2/perlecan). A chimeric molecule with multiple  
 CC domains homologous to the low density lipoprotein receptor, laminin,  
 CC neural cell adhesion molecules, and epidermal growth factor.";  
 CC RT J. Biol. Chem. 267:8544-8557(1992).  
 CC [3]  
 CC RP SEQUENCE OF 1018-1472 FROM N.A.  
 CC TISSUE=Colon;  
 CC MEDLINE=91365376; PubMed=1679749;  
 CC Dodge G.R., Kovalsky I., Chu M.L., Hassell J.R., McBride O.W.,  
 CC Yi H.F., Iozzo R.V.;  
 CC "Heparan sulfate proteoglycan of human colon: partial molecular  
 CC cloning, cellular expression, and mapping of the gene (HSPG2) to the  
 CC short arm of human chromosome 1.";  
 CC RT Genomics 10:673-680(1991).  
 CC [4]  
 CC RP SEQUENCE OF 892-1398 FROM N.A.  
 CC TISSUE=Fibroblasts;  
 CC MEDLINE=92120660; PubMed=1685141;  
 CC Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,  
 CC Trygvaason K.;  
 CC "Cloning of human heparan sulfate proteoglycan core protein,  
 CC assignment of the gene (HSPG2) to 1p36.1-p35 and identification of  
 CC a BamHI restriction fragment length polymorphism.";  
 CC RT Genomics 11:389-396(1991).  
 CC [5]  
 CC RP SEQUENCE OF 1-21 FROM N.A.  
 CC MEDLINE=94052171; PubMed=8234307;  
 CC Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;  
 CC "Structural characterization of the complete human perlecan gene and  
 CC its promoter.";  
 CC RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).  
 CC -1- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT  
 CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC  
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE DIFFERENTIATION



Db 2542 SSSASLIANGHT-----LDLN-----CLVASQAPHT---ITWYKRGK--SLPSPHQIVGSR 2586  
 QY 87 VCVSISSENDNGISFTCRU-----GRDQSVSV-----VLNVFPLPLSGNDPOTVE 134  
 Db 2587 LRIPOVTADSG-EYVCHVSNAGSRETSILVITOGSSSHVPRVS-PIRIESSSTV 2644  
 QY 135 EGSNKLVCNKNKANKPQAVMMYKNSLLDLKSKRHQIVQTSFQSLITVKEPQNGTYS 194  
 Db 2645 ECGTLDLKNVVAROPQALITWYKRGSL--FSRHQ-----FHSGHLRLHWSVMSGSEYV 2697  
 QY 195 CIKSSLKTESLDFHLIYK-DKTVGVPLEP 223  
 Db 2698 CRANNVI--DALEASIVISVSPSGSPAP 2725

RESULT 15  
 DCC\_MOUSE STANDARD; PRT: 1447 AA.  
 AC P70211:  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Tumor suppressor protein DCC precursor.  
 DCC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus  
 OX NCBI Taxid:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-Brain;  
 RX MEDLINE=96112625; PubMed=6570174;  
 RA Cooper H.M., Ames P., Britto J., Gad J., Wilks A.F.:  
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer  
 RT gene (mxc) and its expression in the developing mouse embryo.";  
 RL Oncogene 11:2243-2254(1995).  
 RN [2]  
 RP REVISIONS  
 RC STRAIN-BALB/C; TISSUE-Brain;  
 RA Cooper H.M.:  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF THE PROTEIN ARE PRODUCED FROM  
 CC THE SAME GENE BY THE USE OF ALTERNATIVE INITIATION SITES. A THIKU  
 CC FORM WHICH IS EXPRESSED ONLY IN THE EMBRYO IS PRODUCED BY  
 CC ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE  
 CC DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN  
 CC BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.  
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS  
 CC EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION  
 CC AND REMAIN AT THIS LEVEL IN THE ADULT.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC  
 CC SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

NP InterPro: IPR001600; Iq\_111e.  
 DR Pfam: PF00041; In3; 6.  
 DR Pfam: PF00047; Iq; 4.  
 DR PRINTS: PR00014; FNTYPEIII.  
 DR SMART: SM00060; FN3; 6.  
 DR SMART: SM00410; IG\_Like; 2.  
 DR SMART: SM00408; IGC2; 3.  
 KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;  
 KW Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1447  
 FT FT 85 1447  
 FT CHAIN 85 1447  
 FT FT 85 1447  
 FT INIT\_MET 85 85  
 FT DOMAIN 26 1097  
 FT TRANSMEM 1098 1122  
 FT DOMAIN 1123 1447  
 FT DOMAIN 54 124  
 FT DOMAIN 154 219  
 FT DOMAIN 254 317  
 FT DOMAIN 345 407  
 FT DOMAIN 426 522  
 FT DOMAIN 525 618  
 FT DOMAIN 619 716  
 FT DOMAIN 722 816  
 FT DOMAIN 840 940  
 FT DOMAIN 941 1042  
 FT DISULFID 61 117  
 FT DISULFID 161 212  
 FT DISULFID 261 310  
 FT DISULFID 352 400  
 FT CARBOHYD 60 60  
 FT CARBOHYD 94 94  
 FT CARBOHYD 299 299  
 FT CARBOHYD 318 318  
 FT CARBOHYD 478 478  
 FT CARBOHYD 628 628  
 FT CARBOHYD 702 702  
 FT VANSPLIC 819 838  
 FT SEQUENCE 1447 AA; 158298 MW; 0D1F1097C22D5B9F CR64;

Query Match 11.0%; Score 147; DB 1; Length 1447;  
 Best Local Similarity 25.8%; Pred. No. 0.00041;  
 Matches 51; Conservative 37; Mismatches 86; Indels 24; Gaps 9;

QY 19 ILFLPREMTSSVLTVNGKTEYIIIDTPGSAQLCAVONHTREELIWRREGSDVLS 78  
 Db 240 LVFLQR--PSNVIAIEGK-----DAVLECCVSGYP-PPSFTWLRGFEVQLPS 284  
 QY 79 G--NKINSVSVSSISLSDNLSFTCKLS-KDSYSSVSVLVNTFFP-LLSGNGFIVE 134  
 Db 285 KRYSLGGSNLISVNTDDSG-TTTCVYTKNEIJSASALTYLVLPWPLNHRPSNLAY 343  
 QY 135 EGSNKLVCNKNKANKPQAVMMYKNSLLDLKSKRHQIVQTSFQSLITVKEPQNGTYS 194  
 Db 344 EASMDIEFCVAGSKGVPTVNMKKNQDVV-IPSDYQIVGGS--NLRLGVAKSDEGFYQ 399  
 QY 195 CIKSSLKTESLDFHLIY 212  
 Db 400 CVAENEGNAOSSAQLIV 417

Search completed: September 17, 2002, 13:24:17  
 Job time: 804 sec



GenCore version 4.5  
Copyright (c) 1993 2000 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 08:21:32 : Search time 29:5.1 seconds

(without alignments)  
5472.907 Million cell updates/sec

Title: US-09-863-823-1

Perfect score: 1 acgcaaatgggaagattctt.....acagtgaacagctctatga 765

Sequence:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Scored: 1797656 swgs, 1048326244 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 400000000

Post-processing: Minimum Match 9%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenBank:  
1: gb\_ba:\*  
2: gb\_hhg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_par:\*  
7: gb\_ph:\*  
8: gb\_pi:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vt:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pl:\*  
24: em\_ph:\*  
25: em\_pi:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vt:\*  
30: em\_hhg\_hum:\*  
31: em\_hhg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	504.6	66.0	1340	10	BC022683
2	286.6	37.5	70248	2	AC103795
3	286.6	37.5	171136	3	AC090685
4	286.6	37.5	172046	2	AC087644
5	284	37.1	17444	4	AC04485
6	277.4	36.3	3338	6	AX245713
7	227.2	29.7	362	6	AX246249
8	225.4	29.5	328	5	AX246408
9	223.6	29.2	359	6	AX070657
10	214.6	28.1	201797	2	AL645479
11	214.6	28.1	229289	10	AL63842
12	179.8	23.5	359	6	AX070653
13	109.6	14.3	70248	2	AC103795
14	43.8	5.7	217730	2	AC087440
15	42	5.5	15020	5	AF429315
16	42	5.5	217730	2	AC087440
17	40.2	5.3	393	11	G37560
18	39.6	5.2	2134	3	DM051473
19	39.6	5.2	2134	3	DM051473
20	39.6	5.2	68727	4	AF004516
21	39.6	5.2	194264	3	AC007579
22	39.6	5.2	462173	3	AE003808
23	39.2	5.1	2799	9	HSNCAME
24	39.2	5.1	2940	9	S21824
25	39.2	5.1	3309	9	AK057509
26	39.2	5.1	4092	6	AK062746
27	39.2	5.1	4092	6	HUMKOR4A
28	39.2	5.1	8513	6	AX07284
29	39	5.1	172567	2	AC015493
30	39	5.1	203252	2	AC025421
31	38.8	5.1	157756	2	AP001075
32	38.8	5.1	195622	2	AP00807
33	38.6	5.0	2855	10	RN034985
34	38.6	5.0	2863	6	A72401
35	38.6	5.0	2863	6	AP043363
36	38.6	5.0	2863	6	AP043363
37	38.6	5.0	65119	6	AL560062
38	38.6	5.0	125896	6	HS04628
39	38.6	5.0	159287	9	AP003694
40	38.6	5.0	175216	2	AC027031
41	38.6	5.0	177741	2	AC018937
42	38.4	5.0	291	1	MC385
43	38.2	5.0	118368	9	AC076412
44	38	5.0	71023	2	AC004425
45	38	5.0	106585	9	AL160392

#### ALIGNMENTS

RESULT	1	1440 bp	muscu	1440 bp	muscu
BC022683	Mus musculus, RIKEN cDNA 2010002A20 gene, clone IMAGE:4241194,				
DEFINITION	BC022683				
ACCESSION	BC022683				
VERSION	BC022683.1				
KEYWORDS	house mouse;				
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;				
REFERENCE	1 (bases 1 to 1340)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-EB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>				
COMMENT	Contact: <a href="mailto:cgaps-remail.nih.gov">cgaps-remail.nih.gov</a>				
	Issue Procurement: Jeffrey E. Green, M.D.				



\* NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
688 787: gap of 100 bp in length  
788 1501: contig of 714 bp in length  
1502 1601: gap of 100 bp  
1602 2345: contig of 744 bp in length  
2346 2445: gap of 100 bp  
2446 3151: contig of 706 bp in length  
3152 3251: gap of 100 bp  
3252 3929: contig of 678 bp in length  
3930 4029: gap of 100 bp  
4030 4719: contig of 690 bp in length  
4720 4819: gap of 100 bp  
4820 5538: contig of 719 bp in length  
5539 5638: gap of 100 bp  
5639 6356: contig of 718 bp in length  
6357 6456: gap of 100 bp  
6457 7156: contig of 700 bp in length  
7157 7256: gap of 100 bp  
7257 7888: contig of 632 bp in length  
7889 7988: gap of 100 bp  
7989 8743: contig of 755 bp in length  
8744 8843: gap of 100 bp  
8844 9533: contig of 690 bp in length  
9534 9613: gap of 100 bp  
9634 10371: contig of 738 bp in length  
10372 10471: gap of 100 bp  
10472 11203: contig of 732 bp in length  
11204 11303: gap of 100 bp  
11304 12016: contig of 713 bp in length  
12017 12116: gap of 100 bp  
12117 12823: contig of 707 bp in length  
12824 12923: gap of 100 bp  
12924 13628: contig of 705 bp in length  
13629 13728: gap of 100 bp  
13729 14441: contig of 713 bp in length  
14442 14541: gap of 100 bp  
14542 15272: contig of 731 bp in length  
15273 15372: gap of 100 bp  
15373 16106: contig of 734 bp in length  
16107 16206: gap of 100 bp  
16207 16935: contig of 729 bp in length  
16936 17035: gap of 100 bp  
17036 17762: contig of 727 bp in length  
17763 17862: gap of 100 bp  
17863 18558: contig of 696 bp in length  
18559 18658: gap of 100 bp  
18659 19351: contig of 693 bp in length  
19352 19451: gap of 100 bp  
19452 20181: contig of 730 bp in length  
20182 20281: gap of 100 bp  
20282 21023: contig of 742 bp in length  
21024 21123: gap of 100 bp  
21124 21827: contig of 704 bp in length  
21828 21927: gap of 100 bp  
21929 22624: contig of 697 bp in length  
22625 22724: gap of 100 bp  
22725 23462: contig of 677 bp in length  
23462 23501: gap of 100 bp  
23502 24228: contig of 727 bp in length  
24229 24328: gap of 100 bp  
24329 25055: contig of 727 bp in length

25056 25155: gap of 100 bp  
25156 25898: contig of 743 bp in length  
25899 25994: gap of 100 bp  
25995 26710: contig of 712 bp in length  
26711 26810: gap of 100 bp  
26811 27551: contig of 741 bp in length  
27552 27651: gap of 100 bp  
27652 28394: contig of 743 bp in length  
28395 29230: contig of 736 bp in length  
29231 29330: gap of 100 bp  
29331 30037: contig of 707 bp in length  
30038 30137: gap of 100 bp  
30138 30854: contig of 717 bp in length  
30855 30954: gap of 100 bp  
30955 31680: contig of 726 bp in length  
31681 31780: gap of 100 bp  
31781 32520: contig of 740 bp in length  
32521 32620: gap of 100 bp  
32621 33354: contig of 734 bp in length  
33355 33454: gap of 100 bp  
33455 34191: contig of 737 bp in length  
34192 34291: gap of 100 bp  
34292 35034: contig of 743 bp in length  
35035 35134: gap of 100 bp  
35135 35680: contig of 746 bp in length  
35681 35940: gap of 100 bp  
35941 36663: contig of 683 bp in length  
36664 37477: gap of 100 bp  
37478 37577: contig of 714 bp in length  
37578 38321: contig of 744 bp in length  
38322 38421: gap of 100 bp  
38422 39110: contig of 689 bp in length  
39111 39210: gap of 100 bp  
39211 39892: contig of 682 bp in length  
39893 39992: gap of 100 bp  
39993 40678: contig of 686 bp in length  
40679 40778: gap of 100 bp  
40779 41479: contig of 701 bp in length  
41480 41579: gap of 100 bp  
41580 42110: contig of 731 bp in length  
42111 42410: gap of 100 bp  
42411 43344: contig of 734 bp in length  
43345 43244: gap of 100 bp  
43245 43965: contig of 741 bp in length  
43966 44085: gap of 100 bp  
44086 44777: contig of 692 bp in length  
44778 44877: gap of 100 bp  
44878 45598: contig of 721 bp in length  
45599 45698: gap of 100 bp  
45699 46433: contig of 735 bp in length  
46434 46533: gap of 100 bp  
46534 47267: contig of 734 bp in length  
47268 47367: gap of 100 bp  
47368 48058: contig of 691 bp in length  
48059 48158: gap of 100 bp  
48159 48880: contig of 722 bp in length  
48881 48980: gap of 100 bp  
48981 49717: contig of 777 bp in length  
49718 49817: gap of 100 bp  
49818 50560: contig of 743 bp in length  
50561 50660: gap of 100 bp  
50661 51400: contig of 740 bp in length  
51401 51500: gap of 100 bp  
51501 52227: contig of 727 bp in length  
52228 52327: gap of 100 bp  
52328 53066: contig of 739 bp in length  
53067 53166: gap of 100 bp  
53167 53653: contig of 687 bp in length  
53654 53953: gap of 100 bp  
53954 54675: contig of 722 bp in length  
54676 54775: gap of 100 bp









```

* 24748 24847: gap of unknown length
* 24848 35273: contig of 10426 bp in length
* 35274 35374: gap of unknown length
* 35374 56929: contig of 21556 bp in length
* 56930 57029: gap of unknown length
* 57030 73825: contig of 16796 bp in length
* 73826 73926: gap of unknown length
* 73926 89139: contig of 15213 bp in length
* 89139 89239: gap of unknown length
* 89239 117744: contig of 28505 bp in length
* 117744 117844: gap of unknown length
* 117844 172622: contig of 54779 bp in length.
FEATURES
    source
        1..172622
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="17"
            /clone="RP11-403B9"
            1..6530
                /note="assembly_name:Contig10"
            7031..24747
                /note="assembly_name:Contig11"
            24848..35273
                /note="assembly_name:Contig12"
            35374..56929
                /note="assembly_name:Contig13"
            clone_end:SP6
            vector_side:right"
            57030..73825
                /note="assembly_name:Contig14"
            73926..89138
                /note="assembly_name:Contig15"
            89239..117743
                /note="assembly_name:Contig16"
            117844..172622
                /note="assembly_name:Contig17"
            clone_end:17
            vector_side:right"

```

```

BASE COUNT 49185 a 36666 c 36708 g 49156 t 707 others
ORIGIN

```

```

Query Match          37 1%: Score 284; DB 2; Length 172622;
Best Local Similarity 91.2%: Pred. No. 1..1e-72;
Matches 299; Conservative 2; Mismatches 27; Indels 0; Gaps 0;

OY 12 agattctctctctgtaatttattcttcgcaagtgagatgacaagttctgtttaac 71
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 160044 AGGTTGTTACTCTGTAATCAATAGTTTCCTTACTCTCCAGCTTCTGTTTAA 159985

OY 72 tttgaatgtaaaactggaactatctctgatactactgctcccaagcaactct 131
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159984 TGTGAATGGTAACCTGAGACATATCTCTGGTACTACCTGGCTCCCAAGCATCTCT 159925

OY 132 gatatgtctctcaaacacacacagagagagaaactgtctgttacaagagagag 191
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159924 CATATGTGCTGTTCAAAACACACACAGAGAGAACTGCTGTGTACCAAGAGAG 159865

OY 192 gagaagtgatctgaatctgaaacaaatcaatccagctgtctgtctcttccat 251
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159864 GAGAGTGATTTGAAATCTGGAAACAAATTCATTCACCTGCTGTGTTCTTCCAT 159805

OY 252 caatgaaatgacaacgagatcggttactcctgcaagagctgagagagatcagtcgtgttc 311
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159804 CAGTGAATATGACACAGCAATCAAGCTTTTACTCTGACAGCTGGGAGAGGATCATCCGTGTC 159745

OY 312 cgttcggtggtgctgaatgttacttctt 339
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159744 CGTTTCGTGTGCTGTAATGTACTTGT 159717

```

RESULT 6  
AX245713

```

LOCUS AX245713 338 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 643 from Patent WO0166753.
ACCESSION AX245713
VERSION AX245713.1 GI:15860387
KEYWORDS
SOURCE
    human.
    ORGANISM
        Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
    1 (bases 1 to 338)
    Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
    Kassam,A., Lamson,G., Dmanac,R., Civenjakov,R., Dickson,M.,
    Dmanac,S., Labat,I., Ieshkowitz,D., Kita,D., Garcia,V. and
    Stache-Crain,B.
    Human genes and gene expression products
    Patent: WO 0166753-A 643 13-SEP-2001;
    Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES
    source
        1..338
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            69 c 91 g 85 t
BASE COUNT 93 a 69 c 91 g 85 t
ORIGIN

```

```

Query Match          36.3%: Score 277.4; DB 6; Length 338;
Best Local Similarity 98.2%: Pred. No. 7e-71;
Matches 278; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 55 acaagttctgtttaactgtgaatgtaaaactggaactatatactgatactacact 114
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 56 AGAGCTTCTGTTTAACTGGAATGTTAAACTGAAATATATCTGTGATTAATCTACCT 115

OY 115 gctcccaagcatctgataatgtctgttcaaacacacacagagagaaactgtctc 174
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 GCGTCCCAACCATCTCGATATGCTGCTTCAAAACACACAGAGAGAAATCTGCTC 175

OY 175 tggatccggagagagagagagagagagagagagagagagagagagagagag 234
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 176 TGGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCT 235

OY 235 gctcgtgctctctctctctctctctctctctctctctctctctctctctctct 294
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 236 GCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 295

OY 295 agagatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 296 AGGATCAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 338

```

```

RESULT 7
AX246249 361 bp DNA linear PAT 28-SEP-2001
LOCUS AX246249
DEFINITION Sequence 1179 from Patent WO0166753.
ACCESSION AX246249
VERSION AX246249.1 GI:15860923
KEYWORDS
SOURCE
    human.
    ORGANISM
        Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
    1 (bases 1 to 361)
    Sudduth-Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
    Kassam,A., Lamson,G., Dmanac,R., Civenjakov,R., Dickson,M.,
    Dmanac,S., Labat,I., Ieshkowitz,D., Kita,D., Garcia,V. and
    Stache-Crain,B.
    Human genes and gene expression products
    Patent: WO 0166753-A 1179 13-SEP-2001;
    Chiron Corporation (US) ; Hyseq Inc. (US)
FEATURES
    source
        1..361
            Location/Qualifiers

```

FASTA COUNT 114 a 64 c 90 q 94 l  
 ORIGIN  
 Query Match 29.7%, Score 227.6, DB 6, Length 361,  
 Best Local Similarity 99.1%, Pred. No. 4, He=56;  
 Matches 223; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1 atgcaaatggaagattctctcttaagattatcttgcacgagatgaagaat 60  
 133 ATGCAAAATGGGAAGATTCTCTCTTAAGATTATTTCTGACGATGACAAAGT 192  
 194 TCTGTTTAAGTGTGAATGGTAAGATGACAAATATCTGACATGACAGCTGCTG 252  
 121 caagatctctgaatagtgctggttcaaacacacacagagagagagagatgctg 180  
 253 CAAGCATCTCTGATATGCTGCTGTTCAAAATCACACACAGACAGCAAGACTGCTGCTG 312  
 181 caag 228  
 413 CCAAGACAGGAG 360

RESULT 8  
 LOCUS AX246408 328 bp DNA Linear PAT 28-SEP-2001  
 DEFINITION Sequence 1338 from patient W00166753.  
 ACCESSION AX246408  
 VERSION AX246408.1 GI:125861052  
 KEYWORDS  
 SOURCE human  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 328)  
 Williams, L.T., Escobedo, J., Inlis, M.A., Garcia, P.D.,  
 Sudduth-Klinger, J., Keshavarz, F., Kennedy, G.C., Pot, D.,  
 Krasum, A., Lamson, G., Dymack, R., Gravenjakov, R., Dickson, M.,  
 Dimauc, S., Labat, L., Leskowitz, D., Kila, D., Garcia, V. and  
 Straché Crain, B.  
 TITLE Human genes and gene expression products  
 JOURNAL Patient: W0 0166753-A 1338 13-SEP-2001;  
 Chiron Corporation (US); Hysq, Inc. (US)  
 FEATURES  
 SOURCE 1..328  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 102 a 62 c 81 q 83 l  
 ORIGIN  
 Query Match 29.5%, Score 225.4, DB 6, Length 328,  
 Best Local Similarity 97.8%, Pred. No. 1,6e-55;  
 Matches 223; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

1 atgcaaatggaagattctctcttaagattatcttgcacgagatgaagaat 60  
 97 ATGCAAAATGGGAAGATTCTCTCTTAAGATTATTTCTGACGATGACAAAGT 156  
 61 tctgtttaaactgtgaatgtaaaactgaactatactctgaactacactgctgctc 120  
 157 TCTGTTTAAGTGTGAATGGTAAGATGACAAATATCTGATGACACTGGCTGCTC 216  
 121 caagatctctgaatagtgctggttcaaacacacacagagagagagagatgctg 180  
 217 CAAGCATCTCTGATATGCTGCTGTTCAAAATCACACACAGACAGCAAGACTGCTGCTG 276  
 181 caag 231  
 411 TCTGTTTAAGTGTGAATGTAAGATGTAAGATGTAAGATGTAAGATGTAAGATG 471

18 277 CCAAGACAGGAG 327  
 RESULT 9  
 LOCUS AX070657 369 bp DNA Linear PAT 25-JAN-2001  
 DEFINITION Sequence 1129 from Patient W00125568.  
 ACCESSION AX070657  
 VERSION AX070657.1 GI:12580442  
 KEYWORDS  
 SOURCE human  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 359)  
 Williams, L.T., Escobedo, J., Inlis, M.A., Garcia, P.D., Klinger, J.,  
 Krasum, A., Lamson, G., Dymack, R., Gravenjakov, R., Dickson, M.,  
 Labat, L., Leskowitz, D., Kila, D., Garcia, V. and Straché Crain, B.  
 TITLE Human genes and gene expression products  
 JOURNAL Patient: W0 0125568-A 1129 11-JAN-2001;  
 Chiron Corporation (US); Hysq, Inc. (US)  
 FEATURES  
 SOURCE 1..359  
 location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

FASTA COUNT 113 a 64 c 92 q 89 t 1 others  
 ORIGIN  
 Query Match 29.2%, Score 227.6; DB 6; Length 359;  
 Best Local Similarity 98.7%; Pred. No. 5.6e-55;  
 Matches 223; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 atgcaaatggaagattctctcttaagattatcttgcacgagatgaagaat 60  
 133 ATGCAAAATGGGAAGATTCTCTCTTAAGATTATTTCTGACGATGACAAAGT 192  
 61 tctgtttaaactgtgaatgtaaaactgaactatactctgaactacactgctgctc 120  
 194 TCTGTTTAAGTGTGAATGGTAAGATGACAAATATCTGACATGACAGCTGCTG 252  
 121 caagatctctgaatagtgctggttcaaacacacacagagagagagagatgctg 180  
 253 CAAGCATCTCTGATATGCTGCTGTTCAAAATCACACACAGACAGCAAGACTGCTGCTG 312  
 181 caag 226  
 413 CCAAGACAGGAG 458

RESULT 10  
 LOCUS AL645479 201707 bp DNA Linear HTG 01-FEB-2002  
 DEFINITION Mus musculus chromosome 11 clone RP23-39603. \*\*\* SEQUENCING IN  
 PECKRESS \*\*\* in unordered pieces.  
 ACCESSION AL645479  
 VERSION AL645479.7 GI:18491374  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTROP.  
 SOURCE mouse  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (sites)  
 McElay, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (31 JAN 2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 hinxton@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk  
 On Feb 4, 2002 this sequence version replaced gi:18476855.  
 COMMENT \*\*\*\*\* Genome Center

```
fragment_chain:4
```

as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following



TITLE  
JOURNAL  
COMMENT

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information  
Center project name: 121770

Center clone name: 2529\_P.2

NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Puns of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 688 787: gap of 687 bp in length  
\* 788 1501: contig of 714 bp in length  
\* 1502 1601: gap of 100 bp  
\* 1602 2345: contig of 744 bp in length  
\* 2346 2445: gap of 100 bp  
\* 2446 3151: contig of 706 bp in length  
\* 3152 3251: gap of 100 bp  
\* 3252 3929: contig of 678 bp in length  
\* 3930 4029: gap of 100 bp  
\* 4030 4719: contig of 690 bp in length  
\* 4720 4819: gap of 100 bp  
\* 4820 5538: contig of 719 bp in length  
\* 5539 5638: gap of 100 bp  
\* 5639 6356: contig of 718 bp in length  
\* 6357 6456: gap of 100 bp  
\* 6457 7156: contig of 700 bp in length  
\* 7157 7256: gap of 100 bp  
\* 7257 7888: contig of 632 bp in length  
\* 7889 7988: gap of 100 bp  
\* 7989 8743: contig of 755 bp in length  
\* 8744 8843: gap of 100 bp  
\* 8844 9533: contig of 690 bp in length  
\* 9534 9633: gap of 100 bp  
\* 9634 10371: contig of 738 bp in length  
\* 10372 10471: gap of 100 bp  
\* 10472 11203: contig of 732 bp in length  
\* 11204 11303: gap of 100 bp  
\* 11304 12016: contig of 713 bp in length  
\* 12017 12116: gap of 100 bp  
\* 12117 12823: contig of 707 bp in length  
\* 12824 12923: gap of 100 bp  
\* 12924 13628: contig of 705 bp in length  
\* 13629 13728: gap of 100 bp  
\* 13729 14441: contig of 713 bp in length  
\* 14442 14541: gap of 100 bp  
\* 14542 15272: contig of 731 bp in length  
\* 15273 15372: gap of 100 bp  
\* 15373 16106: contig of 734 bp in length  
\* 16107 16206: gap of 100 bp  
\* 16207 16935: contig of 725 bp in length  
\* 16936 17035: gap of 100 bp

17036 17762: contig of 727 bp in length  
\* 17763 17862: gap of 100 bp  
\* 17863 18558: contig of 696 bp in length  
\* 18559 18658: gap of 100 bp  
\* 18659 19351: contig of 693 bp in length  
\* 19352 19451: gap of 100 bp  
\* 19452 20181: contig of 730 bp in length  
\* 20182 20281: gap of 100 bp  
\* 20282 21023: contig of 742 bp in length  
\* 21024 21123: gap of 100 bp  
\* 21124 21947: contig of 794 bp in length  
\* 21928 21927: gap of 100 bp  
\* 21928 22624: contig of 697 bp in length  
\* 22625 22724: gap of 100 bp  
\* 22725 23401: contig of 677 bp in length  
\* 23402 23501: gap of 100 bp  
\* 23502 24228: contig of 727 bp in length  
\* 24229 24328: gap of 100 bp  
\* 24329 25055: contig of 727 bp in length  
\* 25056 25155: gap of 100 bp  
\* 25156 25898: contig of 743 bp in length  
\* 25899 25998: gap of 100 bp  
\* 25999 26710: contig of 712 bp in length  
\* 26711 26810: gap of 100 bp  
\* 26811 27551: contig of 741 bp in length  
\* 27552 27651: gap of 100 bp  
\* 27652 28394: contig of 743 bp in length  
\* 28395 28494: gap of 100 bp  
\* 28495 29230: contig of 736 bp in length  
\* 29231 29330: gap of 100 bp  
\* 29331 30037: contig of 707 bp in length  
\* 30038 30137: gap of 100 bp  
\* 30138 30854: contig of 717 bp in length  
\* 30855 30954: gap of 100 bp  
\* 30955 31680: contig of 726 bp in length  
\* 31681 31780: gap of 100 bp  
\* 31781 32520: contig of 740 bp in length  
\* 32521 32620: gap of 100 bp  
\* 32621 33354: contig of 734 bp in length  
\* 33355 33454: gap of 100 bp  
\* 33455 34191: contig of 737 bp in length  
\* 34192 34291: gap of 100 bp  
\* 34292 35034: contig of 743 bp in length  
\* 35035 35134: gap of 100 bp  
\* 35135 35880: contig of 746 bp in length  
\* 35881 35980: gap of 100 bp  
\* 35981 36663: contig of 683 bp in length  
\* 36664 36763: gap of 100 bp  
\* 36764 37477: contig of 714 bp in length  
\* 37478 37577: gap of 100 bp  
\* 37578 38321: contig of 744 bp in length  
\* 38322 38421: gap of 100 bp  
\* 38422 39110: contig of 689 bp in length  
\* 39111 39210: gap of 100 bp  
\* 39211 39892: contig of 682 bp in length  
\* 39893 39992: gap of 100 bp  
\* 39993 40678: contig of 686 bp in length  
\* 40679 40778: gap of 100 bp  
\* 40779 41479: contig of 701 bp in length  
\* 41480 41579: gap of 100 bp  
\* 41580 42310: contig of 731 bp in length  
\* 42311 42410: gap of 100 bp  
\* 42411 43144: contig of 734 bp in length  
\* 43145 43244: gap of 100 bp  
\* 43245 43985: contig of 741 bp in length  
\* 43986 44085: gap of 100 bp  
\* 44086 44777: contig of 692 bp in length  
\* 44778 44877: gap of 100 bp  
\* 44878 45598: contig of 721 bp in length  
\* 45599 45698: gap of 100 bp  
\* 45699 46433: contig of 735 bp in length  
\* 46434 46533: gap of 100 bp  
\* 46534 47267: contig of 734 bp in length





```

* 197288 197387: gap of 100 bp
* 197388 198887: contig of 1500 bp in length
* 198888 198987: gap of 100 bp
* 198988 200386: contig of 1399 bp in length
* 200387 200486: gap of 100 bp
* 200487 201763: contig of 1277 bp in length
* 201764 201863: gap of 100 bp
* 201864 203900: contig of 2037 bp in length
* 203901 204000: gap of 100 bp
* 204001 205876: contig of 1876 bp in length
* 205877 205976: gap of 100 bp
* 205977 207807: contig of 1831 bp in length
* 207808 207907: gap of 100 bp
* 207908 210489: contig of 2582 bp in length
* 210490 210589: gap of 100 bp
* 210590 213765: contig of 3176 bp in length
* 213766 213865: gap of 100 bp
* 213866 217087: contig of 3222 bp in length
* 217088 217187: gap of 100 bp
* 217188 217730: contig of 543 bp in length.

```

## FEATURES

```

source
1. 217730
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-425G13"
/clone_id="RP11 Human Male BAC"
1. 182945
/note="assembly_fragment"
vector_side:left"
183050..183137
/note="assembly_fragment"
183238..183640
/note="assembly_fragment"
183741..184713
/note="assembly_fragment"
184814..185319
/note="assembly_fragment"
185420..186040
/note="assembly_fragment"
186141..186820
/note="assembly_fragment"
186921..187855
/note="assembly_fragment"
187956..188593
/note="assembly_fragment"
188694..189289
/note="assembly_fragment"
189390..190844
/note="assembly_fragment"
190923..191663
/note="assembly_fragment"
191764..192866
/note="assembly_fragment"
192967..194120
/note="assembly_fragment"
194221..194904
/note="assembly_fragment"
195005..195907
/note="assembly_fragment"
196008..197287
/note="assembly_fragment"
197388..198887
/note="assembly_fragment"
198988..200386
/note="assembly_fragment"
200487..201763
/note="assembly_fragment"
201864..203900
/note="assembly_fragment"
204001..205876

```

```

/note="assembly_fragment"
205977..207807
/note="assembly_fragment"
207908..210489
/note="assembly_fragment"
210590..213765
/note="assembly_fragment"
213866..217087
/note="assembly_fragment"
217188..217730
/note="assembly_fragment"
clone_end:177
vector_side:right"
BASE COUNT 69080 a 40450 c 39974 g 65624 t 2602 others
ORIGIN
Query Match 5.7% Score 43.8; DB 2; Length 217730;
Best Local Similarity 95.7% Pred. No. 0.11;
Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 719 agcttcgcatgaagataaagaccctcacagtggaacagctctatga 765
|||||
ub 199852 acctctcagatgaagatvaatgaacccctcacagtcagctctgttaa 199806
|||||
RESULT 15
AF429315/C AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
TITLE Submitted (05-OCT-2001) psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
source
1. 125020
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3: between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit=ctg
complement(<36507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/note="JPH3"
complement(<36507..36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 2002, 09:59:08 : Search time 29.5.1 Seconds

(without alignments)  
5644.606 Million cell updates/sec

Title: US-09-863-823-5

Perfect score: 789

Sequence: 1 atgtatgaaagagatgagt

atgtatgaaagagatgagt /84

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 356512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_hrg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hrg\_hum: \*  
31: em\_hrg\_inv: \*  
32: em\_hrg\_other: \*  
33: em\_hrgo\_inv: \*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

# SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	504.6	64.0	1340	10	BC022683
2	286.6	36.3	17136	2	AC090685
3	286.6	36.3	172046	2	AC087644
4	286.6	36.3	172046	2	AC024065
5	277.4	35.2	172622	6	AX245713
6	277.4	35.2	172622	6	AX245713
7	250.8	31.8	361	6	AX246249
8	249	31.6	359	6	AX246249
9	247.2	31.4	201797	2	AL645479
10	214.6	27.2	229289	10	AL603842
11	214.6	27.2	229289	10	AL603842
12	203.4	25.8	359	6	AC070653
13	109.6	13.9	70248	4	AC033795
14	43.8	5.4	21734	4	AC087440
15	42	5.3	125020	9	AF429315
16	42	5.3	217730	9	AC087440
17	40.4	5.1	174493	9	AC025522
18	40.4	5.1	189126	2	AC021132
19	40.4	5.1	189126	2	AC021132
20	39.6	5.0	2134	3	DM051473
21	39.6	5.0	8536	2	AC020364
22	39.6	5.0	68727	3	AC004516
23	39.6	5.0	193262	3	AC007679
24	39.6	5.0	262173	3	AE003808
25	39.6	5.0	2799	9	HSNCAME
26	39.6	5.0	2960	9	S71834
27	39.6	5.0	4009	9	AK07509
28	39.6	5.0	4092	6	AP062746
29	39.6	5.0	4092	6	H09082A
30	39.6	5.0	8513	6	AX207284
31	39	4.9	172567	2	AC015493
32	39	4.9	203252	2	AC024421
33	38.8	4.9	157756	2	AP001075
34	38.8	4.9	157756	2	AP001075
35	38.8	4.9	195624	2	AP008007
36	38.6	4.9	2855	10	RN034985
37	38.6	4.9	2869	6	A72401
38	38.6	4.9	2869	6	AR043363
39	38.6	4.9	2869	6	160497
40	38.6	4.9	55113	9	AL590062
41	38.6	4.9	125698	9	HS69228
42	38.6	4.9	159287	9	AP003694
43	38.6	4.9	175218	2	AC027041
44	38.6	4.9	177731	2	AC018537
45	38	4.8	291	1	MC385
		4.8	71023	2	AC004426

## ALIGNMENTS

Result	1	1340 bp	muscle	100% identity	100% identity
BC022683	BC022683	1340 bp	muscle	100% identity	100% identity
DEFINITION	Mus musculus, RIKEN cDNA 2010002A20 gene, clone IMAGE:4241194.				
ACCESSION	BC022683				
VERSION	BC022683.1 GI:18490500				
KEYWORDS	house mouse, Mus musculus				
SOURCE	house mouse, Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Cladoda; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB 2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: <a href="http://mgc.ncl.nih.gov">http://mgc.ncl.nih.gov</a>				
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgabs@mail.nih.gov">cgabs@mail.nih.gov</a> Tissue Procurement: Jeffrey E. Green, M.D.				



\*\*\*\*\* NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
688 787: gap of 100 bp  
788 1501: contig of 714 bp in length  
1502 1601: gap of 100 bp  
1602 2345: contig of 744 bp in length  
2346 2445: gap of 100 bp  
2446 3151: contig of 706 bp in length  
3152 3251: gap of 100 bp  
3252 3929: contig of 678 bp in length  
3930 4029: gap of 100 bp  
4030 4719: contig of 690 bp in length  
4720 4819: gap of 100 bp  
4820 5538: contig of 719 bp in length  
5539 5638: gap of 100 bp  
5639 6356: contig of 718 bp in length  
6357 6456: gap of 100 bp  
6457 7156: contig of 700 bp in length  
7157 7256: gap of 100 bp  
7257 7888: contig of 632 bp in length  
7889 7988: gap of 100 bp  
7989 8743: contig of 755 bp in length  
8744 8843: gap of 100 bp  
8844 9533: contig of 690 bp in length  
9534 9633: gap of 100 bp  
9634 10371: contig of 738 bp in length  
10372 10471: gap of 100 bp  
10472 11203: contig of 732 bp in length  
11204 11303: gap of 100 bp  
11304 12016: contig of 713 bp in length  
12017 12116: gap of 100 bp  
12117 12823: contig of 707 bp in length  
12824 12923: gap of 100 bp  
12924 13628: contig of 705 bp in length  
13629 13728: gap of 100 bp  
13729 14441: contig of 713 bp in length  
14442 14541: gap of 100 bp  
14542 15272: contig of 731 bp in length  
15273 15372: gap of 100 bp  
15373 16106: contig of 734 bp in length  
16107 16206: gap of 100 bp  
16207 16935: contig of 729 bp in length  
16936 17035: gap of 100 bp  
17036 17762: contig of 727 bp in length  
17763 17862: gap of 100 bp  
17863 18558: contig of 696 bp in length  
18559 18658: gap of 100 bp  
18659 19351: contig of 693 bp in length  
19352 19451: gap of 100 bp  
19452 20181: contig of 730 bp in length  
20182 20281: gap of 100 bp  
20282 21023: contig of 742 bp in length  
21024 21123: gap of 100 bp  
21124 21827: contig of 704 bp in length  
21828 21927: gap of 100 bp  
21928 22524: contig of 697 bp in length  
22525 22724: gap of 100 bp  
22725 23401: contig of 677 bp in length  
23402 23501: gap of 100 bp  
23502 24328: contig of 727 bp in length  
24329 24328: gap of 100 bp  
24329 25055: contig of 727 bp in length

25056 25155: gap of 100 bp  
25156 25898: contig of 743 bp in length  
25899 25998: gap of 100 bp  
25999 26710: contig of 712 bp in length  
26711 26810: gap of 100 bp  
26811 27551: contig of 741 bp in length  
27552 27651: gap of 100 bp  
27652 28394: contig of 743 bp in length  
28395 28494: gap of 100 bp  
28495 29236: contig of 736 bp in length  
29237 29336: gap of 100 bp  
29337 30037: contig of 707 bp in length  
30038 30137: gap of 100 bp  
30138 30854: contig of 717 bp in length  
30855 30954: gap of 100 bp  
30955 31680: contig of 726 bp in length  
31681 31780: gap of 100 bp  
31781 32520: contig of 740 bp in length  
32521 32620: gap of 100 bp  
32621 33354: contig of 734 bp in length  
33355 33454: gap of 100 bp  
33455 34191: contig of 737 bp in length  
34192 34291: gap of 100 bp  
34292 35034: contig of 743 bp in length  
35035 35134: gap of 100 bp  
35135 35880: contig of 746 bp in length  
35881 35980: gap of 100 bp  
35981 36661: contig of 683 bp in length  
36662 36761: gap of 100 bp  
36762 37477: contig of 714 bp in length  
37478 37577: gap of 100 bp  
37578 38321: contig of 744 bp in length  
38322 38421: gap of 100 bp  
38422 39110: contig of 689 bp in length  
39111 39210: gap of 100 bp  
39211 39892: contig of 682 bp in length  
39893 39992: gap of 100 bp  
39993 40678: contig of 686 bp in length  
40679 40778: gap of 100 bp  
40779 41479: contig of 701 bp in length  
41480 41579: gap of 100 bp  
41580 42310: contig of 731 bp in length  
42311 42410: gap of 100 bp  
42411 43144: contig of 734 bp in length  
43145 43244: gap of 100 bp  
43245 43985: contig of 741 bp in length  
43986 44085: gap of 100 bp  
44086 44777: contig of 692 bp in length  
44778 44877: gap of 100 bp  
44878 45598: contig of 721 bp in length  
45599 45698: gap of 100 bp  
45699 46433: contig of 735 bp in length  
46434 46533: gap of 100 bp  
46534 47267: contig of 734 bp in length  
47268 47367: gap of 100 bp  
47368 48058: contig of 691 bp in length  
48059 48158: gap of 100 bp  
48159 48880: contig of 722 bp in length  
48881 48980: gap of 100 bp  
48981 49717: contig of 737 bp in length  
49718 49817: gap of 100 bp  
49818 50560: contig of 743 bp in length  
50561 50660: gap of 100 bp  
50662 51400: contig of 740 bp in length  
51401 51500: gap of 100 bp  
51501 52227: contig of 727 bp in length  
52228 52327: gap of 100 bp  
52328 53066: contig of 739 bp in length  
53067 53166: gap of 100 bp  
53167 53853: contig of 687 bp in length  
53854 53953: gap of 100 bp  
53954 54675: contig of 722 bp in length  
54676 54775: gap of 100 bp





```

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gen/index.shtml
Project Information
Center project name: E_NIB040FE09
Summary Statistics
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: dye primer ET; 100% of reads
Chemistry: dye terminator H14 dye; 0% of reads
Assembly program: Phrap; version 0.990119
Consensus quality: 167648 bases at least Q40
Consensus quality: 169424 bases at least Q30
Consensus quality: 176452 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 173251; sum-of-contigs
Quality coverage: 5.19 in Q20 bases; agarose-fp
Quality coverage: 5.86 in Q20 bases; sum-of-contigs
* WMP: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 6930: contig of 6930 bp in length
* 6931 7030: gap of unknown length
* 7031 24747: contig of 17717 bp in length

```



Query Match	36.0%	Score 284	DB 2	Length 172632
Overall Similarity	91.28%	Pred. No. 2.8e-72		
Matches 299	Conservative	2	Mismatches 27	Indels 0
QY 36	aagatctctctcttaagtaatttaattcttcgcacgtagagatgacaagcttggttaac	95		
Db 160044	AGGTTGTTTACTCTGTAATATCTAAATAGTTTCTCCCTTACTCCAGGTTCTGTTTAA	159985		
QY 96	tgtgaatgtaaaacgtgagactatactctgatactaacacctgtgtcccaagcatct	155		
Db 159984	TGTGAATGTGTAACACGACGACTATCTCTGATCTACTACCTGGCTCCCAAGCATCTCT	159925		
QY 156	gatactgctcgtctcaaacacacacccagagaagaataactgtctcgtatccgaagaggg	215		
Db 159924	GATATGTGCTGTTTCCAAACCCACCCGAGAGGAAACACTGCTCTGTGTAACCGAGAGGAGG	159865		
QY 216	gagagtgatcttgaactctgaaacaacaaatgaattccaagctgtcgtgctctccat	275		
Db 159864	GAGAGTGGATTGTAATCTGGAACCAAAATTCATTTCCACCTCTGTCTGTGCTTTCCAT	159805		
QY 276	cagtgaaatagacaacgaaatcagcttcaactcgaagctgggagagatcaagtcctgtc	335		
Db 159804	CAGTGAATAATGACAAACGGAATCTAGCTTTACTCTGCAGAGGCTGGAGGAGATCAGTCCGTC	159745		
QY 336	cgttcgtggtgctgctgaatgtaactt 363			
Db 159744	CGTTTCGTGTGTGTAATGTTACTGTG 159717			

LOCUS	AX245713	338 bp	DNA		linear	PAT 28-SEP-2000					
DEFINITION	Sequence 643 from Patent WO0166753.										
ACCESSION	AX245713										
VERSION	AX245713.1	GI:15860387									
KEYWORDS	human.										
SOURCE	human.										
ORGANISM	Homo sapiens										
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 338) Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Sudduth,Klinger,J., Reinhard,C., Ruchazzo,F., Kennedy,G.C., Pot,D., Kassam,A., Lawson,G., Drmanac,P., Crevenkova P., Dickson,M., Drmanac,S., Labat,I., Lesnikowicz,D., Kita,D., Garcia,V. and Stache-Crain,B.										
AUTHORS	Human genes and gene expression products Patent: WO 0166753-A 643 13-SEP-2001; Chiron Corporation (US); Hyseq Inc. (US)										
TITLE	Location/Qualifiers										
JOURNAL	1..338										
FEATURES	source										
BASE COUNT	/organism="Homo sapiens" /db_xref="taxon:9606"										
ORIGIN	93 a 69 c 91 g 85 t										
Query Match	35.2%; Score 277.4; DB 6; Length 338;										
Best Local Similarity	98.2%; Pred. No. 1.8e-70;										
Matches	278; Conservative 2; Mismatches 3; Indels 0; Gaps 0.										
OY	79	acaagttctgtttaaactgtagaaagglaaacggaacatcatcctgatctacacct	138								
DB	56	ACAAGTTCTGTTTAACTGTCAATTCATAAATTAGAAGCATATATCTCGATTAACACT	115								
OY	139	ggctcccaagacatcttgatatgtgtgttcacaaaccacacagaaggaagactgtctc	198								
DB	116	GGCCTCCCAAGCATCTCTGATATGTCTGTTTCAAACCACACAGAGGAAGAACTGGCTC	175								
OY	175	tgttaacagaaaggaaggaagtgtatttgcatttgcataatctgcataaatctacatct	258								
DB	176	TGCTAACGAGAGAGAGAGAGAGAGATTCGAATCTGGAAACAAAATCAATTCACACTCT	235								
OY	259	gtctgtgtctctctctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	318								
DB	295	GTTGTGTGTCT	295								
OY	319	agggaatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgat	361								
DB	296	AGGGATCAGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCGTGCTT	338								
RESULT	7										
LOCUS	AX246249	361 bp	DNA		linear	PAT 28-SEP-2001					
DEFINITION	Sequence 1179 from Patent WO0166753.										
ACCESSION	AX246249										
VERSION	AX246249.1	GI:15860923									
KEYWORDS	human.										
SOURCE	human.										
ORGANISM	Homo sapiens										
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 361) Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Sudduth,Klinger,J., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D., Kassam,A., Lawson,G., Drmanac,P., Ctkvenjakov,R., Dickson,M., Drmanac,S., Labat,I., Lesnikowitz,D., Kita,D., Garcia,V. and Stache-Crain,B.										
AUTHORS	Human genes and gene expression products Patent: WO 0166753-A 1179 13-SEP-2001; Chiron Corporation (US); Hyseq Inc. (US)										
TITLE	Location/Qualifiers										
JOURNAL	1..361										
FEATURES	source										

KEYWORDS: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

```

/note="assembly_
fragment_chain:4

```

COMMENT On Oct 30, 2001 this sequence version replaced gi:16304764.  
hmmquery@sanger.ac.uk clone requests. clonerequest@sanger.ac.uk

COMMENT On Oct 30, 2001 this sequence version replaced g1:16304764.  
During sequence assembly data is compared from overlapping clones.

REFERENCE  
2 (bases 1 to 10248)

AUTHORS  
Ritten, B., Jinton, J., Nussbaum, C., Lander, E., Ali, A., Allen, N.,

TITLE  
JOURNAL  
COMMENT

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Chapel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,  
Cooke, P., Deatello, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S.,  
Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, M., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., Lakoque, K.,  
Lamatzos, R., Landers, T., Lehoucq, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Melnick, J.,  
Meneu, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Hern, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phukhian, P., Pierre, N., Pollard, V.,  
Raymond, C., Rella, P., Rieback, M., Riley, P., Rise, C., Poyrov, P.,  
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.sq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L21770

Center clone name: 2529\_0\_2

\* NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
\* 688 787: gap of 100 bp  
\* 788 1501: contig of 714 bp in length  
\* 1502 1601: gap of 100 bp  
\* 1602 2345: contig of 744 bp in length  
\* 2346 2445: gap of 100 bp  
\* 2446 3151: contig of 706 bp in length  
\* 3152 3251: gap of 100 bp  
\* 3252 3929: contig of 678 bp in length  
\* 3930 4029: gap of 100 bp  
\* 4030 4719: contig of 690 bp in length  
\* 4720 4819: gap of 100 bp  
\* 4820 5538: contig of 719 bp in length  
\* 5539 5638: gap of 100 bp  
\* 5639 6356: contig of 718 bp in length  
\* 6357 6456: gap of 100 bp  
\* 6457 7156: contig of 700 bp in length  
\* 7157 7256: gap of 100 bp  
\* 7257 7888: contig of 632 bp in length  
\* 7889 7988: gap of 100 bp  
\* 7989 8743: contig of 755 bp in length  
\* 8744 8843: gap of 100 bp  
\* 8844 9533: contig of 690 bp in length  
\* 9534 9633: gap of 100 bp  
\* 9634 10371: contig of 738 bp in length  
\* 10372 10471: gap of 100 bp

10472 11203: contig of 732 bp in length  
\* 11204 11303: gap of 100 bp  
\* 11304 12016: contig of 713 bp in length  
\* 12017 12116: gap of 100 bp  
\* 12117 12823: contig of 707 bp in length  
\* 12824 12923: gap of 100 bp  
\* 12924 13628: contig of 705 bp in length  
\* 13629 13728: gap of 100 bp  
\* 13729 14441: contig of 713 bp in length  
\* 14442 14541: gap of 100 bp  
\* 14542 15272: contig of 731 bp in length  
\* 15273 15372: gap of 100 bp  
\* 15373 16107: contig of 734 bp in length  
\* 16107 16207: gap of 100 bp  
\* 16207 16936: contig of 729 bp in length  
\* 16936 17036: gap of 100 bp  
\* 17036 17762: contig of 727 bp in length  
\* 17763 17862: gap of 100 bp  
\* 17863 18558: contig of 686 bp in length  
\* 18559 18658: gap of 100 bp  
\* 18659 19451: contig of 693 bp in length  
\* 19452 19451: gap of 100 bp  
\* 19452 20181: contig of 730 bp in length  
\* 20182 21023: contig of 742 bp in length  
\* 21024 21123: gap of 100 bp  
\* 21124 21827: contig of 704 bp in length  
\* 21828 21927: gap of 100 bp  
\* 21928 22644: contig of 697 bp in length  
\* 22645 22724: gap of 100 bp  
\* 22725 23491: contig of 677 bp in length  
\* 23492 24501: gap of 100 bp  
\* 24502 24528: contig of 727 bp in length  
\* 24529 24528: gap of 100 bp  
\* 24529 25055: contig of 727 bp in length  
\* 25056 25155: gap of 100 bp  
\* 25156 25898: contig of 743 bp in length  
\* 25899 25998: gap of 100 bp  
\* 25999 26710: contig of 712 bp in length  
\* 26711 26810: gap of 100 bp  
\* 26811 27551: contig of 741 bp in length  
\* 27552 27651: gap of 100 bp  
\* 27652 28394: contig of 743 bp in length  
\* 28395 28494: gap of 100 bp  
\* 28495 29231: contig of 716 bp in length  
\* 29231 29330: gap of 100 bp  
\* 29331 30037: contig of 707 bp in length  
\* 30038 30137: gap of 100 bp  
\* 30138 30854: contig of 717 bp in length  
\* 30855 30954: gap of 100 bp  
\* 30955 31680: contig of 726 bp in length  
\* 31681 31780: gap of 100 bp  
\* 31781 32520: contig of 740 bp in length  
\* 32521 32620: gap of 100 bp  
\* 32621 33354: contig of 734 bp in length  
\* 33355 33454: gap of 100 bp  
\* 33455 34191: contig of 737 bp in length  
\* 34192 34291: gap of 100 bp  
\* 34292 35034: contig of 743 bp in length  
\* 35035 35134: gap of 100 bp  
\* 35135 35880: contig of 746 bp in length  
\* 35881 35980: gap of 100 bp  
\* 35981 36663: contig of 683 bp in length  
\* 36664 37477: gap of 100 bp  
\* 37478 37577: contig of 714 bp in length  
\* 37578 38321: gap of 100 bp  
\* 38322 38421: contig of 744 bp in length  
\* 38422 39110: contig of 689 bp in length  
\* 39111 39210: gap of 100 bp  
\* 39211 39892: contig of 682 bp in length  
\* 39893 39992: gap of 100 bp  
\* 39993 40678: contig of 686 bp in length



```

* 188594 188693: gap of 100 bp
* 188694 189285: contig of 596 bp in length
* 189290 189389: gap of 100 bp
* 189390 190822: contig of 1433 bp in length
* 190823 190922: gap of 100 bp
* 190923 191663: contig of 741 bp in length
* 191664 191763: gap of 100 bp
* 191764 192866: contig of 1103 bp in length
* 192867 192966: gap of 100 bp
* 192967 194120: contig of 1154 bp in length
* 194121 194220: gap of 100 bp
* 194221 194904: contig of 684 bp in length
* 194905 195004: gap of 100 bp
* 195005 195907: contig of 903 bp in length
* 195908 196007: gap of 100 bp
* 196008 197287: contig of 1280 bp in length
* 197288 197387: gap of 100 bp
* 197388 198887: contig of 1500 bp in length
* 198888 198987: gap of 100 bp
* 198988 200386: contig of 1399 bp in length
* 200387 200486: gap of 100 bp
* 200487 201763: contig of 1277 bp in length
* 201764 201863: gap of 100 bp
* 201864 203900: contig of 2037 bp in length
* 203901 204000: gap of 100 bp
* 204001 205876: contig of 1876 bp in length
* 205877 205976: gap of 100 bp
* 205977 207807: contig of 1831 bp in length
* 207808 207907: gap of 100 bp
* 207908 210489: contig of 2582 bp in length
* 210490 210589: gap of 100 bp
* 210590 213765: contig of 3176 bp in length
* 213766 213865: gap of 100 bp
* 213866 217087: contig of 3222 bp in length
* 217088 217187: gap of 100 bp
* 217188 217730: contig of 543 bp in length.

```

## FEATURES

source

```

1. 217730
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone_id="RP11-425G13"
/clone_lib="RPC1-11 Human Male BAC"
1. 182949
/feature="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
183238..183640
/feature="assembly_fragment"
183741..184713
/feature="assembly_fragment"
/feature="assembly_fragment"
184814..185319
/feature="assembly_fragment"
185420..186040
/feature="assembly_fragment"
186141..186820
/feature="assembly_fragment"
186921..187855
/feature="assembly_fragment"
187956..188593
/feature="assembly_fragment"
188694..189289
/feature="assembly_fragment"
189390..190822
/feature="assembly_fragment"
190923..191663
/feature="assembly_fragment"
191764..192866
/feature="assembly_fragment"
192967..194120
/feature="assembly_fragment"

```

```

misc_feature
194221..194904
/feature="assembly_fragment"
195005..195907
/feature="assembly_fragment"
196008..197287
/feature="assembly_fragment"
197388..198887
/feature="assembly_fragment"
198988..200386
/feature="assembly_fragment"
200487..201763
/feature="assembly_fragment"
201864..203900
/feature="assembly_fragment"
204001..205876
/feature="assembly_fragment"
205977..207807
/feature="assembly_fragment"
207908..210489
/feature="assembly_fragment"
210590..213765
/feature="assembly_fragment"
213866..217087
/feature="assembly_fragment"
217188..217730
/feature="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature
BASE COUNT 69080 a 40450 c 39974 g 65624 t 2602 others
ORIGIN

```

```

Query Match 5.64; Score 43.8; DB 2; Length 217730;
Best Local Similarity 95.7%; Pred. No. 0.12;
Matches 45; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 743 agctctgcatgaagataaagaccctcacagctgaacagctctcatga 789
|||||tgcattgacgacgattaaacaccttcacactgaacacacttcgtat 199806
DB 199852 agcttgcattgacgacgattaaacaccttcacactgaacacacttcgtat 199806

```

```

RESULT 15 AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS AF429315
DEFINITION Homo sapiens junctophilin-3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 125020)
Holmes,S.E., O'Hean,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL
Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE
21583737
PUBMED
11694876
2 (bases 1 to 125020)
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 603 N. Wolfe St., Baltimore, MD 21287, USA
JOURNAL
location/Qualifiers
1. 125020
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S540 and WI-12410"

```

## FEATURES

source

/note: \*Isolated from a patient with Huntington's Disease. Like 2 (HD12) \* complement (45581, .35746)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 17, 2002, 10:01:13 Search time 86.46 seconds  
(without alignments)  
326.310 Million cell updates/sec

Title: US-09-863-823-2  
Sequence: 1295  
1 MCMGPFLLVLEFLPEMST.....PKIKKLCMKNDPHSETAL 254

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: A:Gcgcgata/132802.\*  
1: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgcgata/gcgcgata/gcgcgata-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	311	24.0	66	22	AAU18000 Human Immunoglobulin
2	171	13.2	1571	19	AAW42087 Human Down syndrome
3	170	13.1	1510	19	AAW42086 Human Down syndrome
4	169.5	13.1	359	22	AAW42086 Human Down syndrome
5	169.5	13.1	431	20	AAW42086 Human Down syndrome
6	169.5	13.1	431	20	AAW42086 Human Down syndrome
7	166	12.8	848	21	AAW42086 Human Down syndrome
8	162	12.5	404	22	AAW42086 Human Down syndrome
9	158	12.4	1079	22	AAW42086 Human Down syndrome
10	151	12.2	444	20	AAW42086 Human Down syndrome
11	158	12.2	444	22	AAW42086 Human Down syndrome

12	157.5	12.2	1415	22	AAW50692 C. elegans UNC-40
13	157	12.1	467	22	AAW50692 Drosophila melanog
14	155	12.0	595	22	AAW50692 Human secreted pro
15	154.5	11.9	537	22	AAW50692 Human secreted pro
16	154.5	11.9	576	22	AAW50692 Human secreted pro
17	154.5	11.9	869	22	AAW50692 Human secreted pro
18	154.5	11.9	869	22	AAW50692 Human secreted pro
19	154	11.9	582	17	AAW50692 Human secreted pro
20	154	11.9	761	17	AAW50692 Human secreted pro
21	154	11.9	868	18	AAW50692 Human secreted pro
22	154	11.9	868	18	AAW50692 Human secreted pro
23	153	11.8	467	19	AAW50692 Human secreted pro
24	153	11.8	467	21	AAW50692 Human secreted pro
25	153	11.8	467	21	AAW50692 Human secreted pro
26	153	11.8	467	22	AAW50692 Human secreted pro
27	153	11.8	467	22	AAW50692 Human secreted pro
28	153	11.8	467	22	AAW50692 Human secreted pro
29	153	11.8	467	22	AAW50692 Human secreted pro
30	153	11.8	467	22	AAW50692 Human secreted pro
31	152	11.7	322	22	AAW50692 Human secreted pro
32	152	11.7	869	18	AAW50692 Human secreted pro
33	152	11.7	869	18	AAW50692 Human secreted pro
34	151.5	11.7	869	18	AAW50692 Human secreted pro
35	149	11.5	478	17	AAW50692 Human secreted pro
36	149	11.5	860	17	AAW50692 Human secreted pro
37	149	11.5	860	17	AAW50692 Human secreted pro
38	149	11.5	860	17	AAW50692 Human secreted pro
39	149	11.5	1447	16	AAW50692 Human secreted pro
40	149	11.5	1447	22	AAW50692 Human secreted pro
41	149	11.5	1447	22	AAW50692 Human secreted pro
42	149	11.5	2037	22	AAW50692 Human secreted pro
43	148.5	11.5	245	19	AAW50692 Human secreted pro
44	148.5	11.5	308	17	AAW50692 Human secreted pro
45	148.5	11.5	325	17	AAW50692 Human secreted pro

## ALIGNMENTS

RESULT 1  
AAU18000 standard; Protein: 66 AA.

AAU18000:

07-NOV-2001 (first entry)

Human immunoglobulin polypeptide SEQ ID No 145.

Immunoglobulin; signal transduction pathway protein; cancer;  
antiscense therapy; gene therapy; neurological disorder; renal disorder;  
cardiovascular disorder; gastrointestinal disorder; pulmonary disorder;  
reproductive disorder; immune system disorder; proliferative disorder;  
muscular disorder.

Homo sapiens.

WO200155315-A2.

02-AUG-2001.

17-JAN-2001; ZWU1W0-USU1326.

31-JAN-2000; ZWU1W0-USU1326.

04-FEB-2000; ZWU1W0-USU1326.

24-FEB-2000; ZWU1W0-USU1326.

02-MAR-2000; ZWU1W0-USU1326.

16-MAR-2000; ZWU1W0-USU1326.

17-MAR-2000; ZWU1W0-USU1326.

18-APR-2000; ZWU1W0-USU1326.

19-MAY-2000; ZWU1W0-USU1326.

28-JUN-2000; ZWU1W0-USU1326.







XX (PEKE ) PE CORP NY  
 XX  
 XX Venter JC, Adams M, LI PMD, Myers EW:  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB: ABL08023.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure: SEQ ID NO 18552; 21pp + Sequence Listing: English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABB57737-ABB72072).  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pt\_sequences.  
 CC  
 XX Sequence 359 AA;  
 SQ

Query Match 13.1%; Score 169.5; DB 22; Length 359;  
 Best Local Similarity 27.1%; Pred. No. 1.4e-06;

Matches 56; Conservative 42; Mismatches 78; Indels 31; Gaps 10;

QY 8 LVIHLEPREMTSSVLTNGKTEYILDTTPQSASLCAVQNHTEELLWYR-EEGRV 66  
 DB 14 LIAITVGTGLAQTPLISYLTQEQIKDI-9GTVEFACSVG-YAKEYNVILFKITSDPV 70  
 QY 67 DKSGNKI-----NSSS--VCYSSISENDGISTFCR--LGRQSVSVSVYL 109  
 DB 71 LLSGTCTLVKIDSLFSLYDPSSTLYKIQIKGLTGLDAG-LYTCQVSVSVTHKVAEVL 129  
 QY 110 NTFPEPLSGNDFQTV--EEGSNVRLVCNVKANPQQAQMMWYK-NSSLDEKSRHQIQOT 166  
 DB 130 SVTRPVISDNTSGVVASGEVQMEYASGYPTPLTWREMAIPIPTSA-----t 183  
 QY 167 SESFOLSTIKVEKPDNGTYSCIAKSSL 193  
 DB 184 YVGNLIRIKSVKKEDIGLYCVADNGV 210

RESULT 5  
 AAM94996  
 ID AAM94996 standard; Protein: 431 AA.  
 XX  
 AC AAM94996;  
 XX  
 DT 12-MAY-1999 (first entry)  
 XX  
 DE SAF-2 polypeptide (derived from EST).  
 XX  
 XX Stialoadhesin; SAF-2; cellular interaction molecule; cancer; inflammation;  
 KW autoimmunity; allergy; asthma; rheumatoid arthritis; multiple sclerosis;  
 KW cerebellar degeneration; Alzheimer's disease; Parkinson's disease; AIDS;  
 KW amyotrophic lateral sclerosis; head injury; neurological abnormality;  
 KW septic shock; stroke; osteoporosis; osteoarthritis; ischemia reperfusion;  
 KW cardiovascular disease; kidney; myocardial; hypotension; hypertension;  
 KW myelodysplastic; anaemia; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN CA2226910-A.  
 XX  
 PD 02-OCT-1998.

XX 26-MAR-1998; 98CA-2226910.  
 XX  
 XX 02-APR-1997; 97US-0041886.  
 XX  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Erickson-Miller CL, Kikly KK:  
 PT WPI: 1999-106564/10.  
 DR  
 DR  
 XX  
 PT New stialoadhesin (SAF-2) polypeptide and polynucleotide - useful as  
 PT diagnostic reagents and for prevention and treatment of cancer,  
 PT AIDS, and Alzheimer's disease  
 XX  
 PS Claim 14; page 23-24; 38pp: English.  
 XX  
 CC The invention relates to a stialoadhesin polypeptide (SAF-2) which is  
 CC utilised as a cellular interaction molecule. SAF-2 polynucleotides and for  
 CC polypeptides are useful for diagnosing susceptibility to diseases and for  
 CC screening for antagonists, agonists. They can also be used to treat  
 CC diseases associated with a SAF-2 imbalance, including inflammation,  
 CC autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,  
 CC cerebellar degeneration, Alzheimer's disease, Parkinson's disease,  
 CC multiple sclerosis, amyotrophic lateral sclerosis, head injury damage  
 CC and other neurological abnormalities, septic shock, cancer, stroke,  
 CC osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,  
 CC cardiovascular diseases, myocardial infarction, AIDS, hypotension,  
 CC hypertension, myelodysplastic syndromes and other abnormalities, aplastic  
 CC anaemia, male pattern baldness, and bacterial, fungal, protozoan and  
 CC viral infections. SAF-2 polypeptides can be used as vaccines and in gene  
 CC therapy to prevent the diseases. SAF 2 polypeptides are useful for  
 CC mapping genes to chromosomes, allowing gene inheritance to be studied  
 CC through linkage analysis.  
 CC  
 XX Sequence 431 AA;  
 SQ

Query Match 13.1%; Score 169.5; DB 20; Length 431;  
 Best Local Similarity 23.0%; Pred. No. 1.8e-06;

Matches 63; Conservative 53; Mismatches 92; Indels 59; Gaps 11;

QY 16 KEMISSVLVNGKTEYILDTTPQSASLCAVQNHTEELLWYR-EEGRV 72  
 DB 144 KGLSVFVTAITHPDIIIGTLESQHSNLTCSVP-----WACKQGPMTSWIGA 194  
 QY 73 KINS-----SSVCVSSISENDGISTFCR--LGRQSVSVSVYL 115  
 DB 195 SVSPPTARSVLTLPKPDHGTSLTQGLTPGVLTSTVTRVDSYPPNLTMTV 254  
 QY 116 -----LISGNDFOYVEGSSNVKLYCNVKANPQQAQMMWYK-NSSLDEKSRHQIQOT 166  
 DB 225 IGGDATASTALQNGSSISLVGQSILRVCAVSNPARIATWRS--ILICPSF-----s 307  
 QY 167 SESFOLSTIKVEKPDNGTYSCIAKSSLKTESLDFELIWKDTCV--PIEPIAACV--- 221  
 DB 308 SPPYLLLPVAVHVDGEFTCAQHQYQHSISISLQNGEQLSTFVSQYLLAAVQGA 367  
 QY 222 ----VIFTLFCGLIAKKKIKMLQMK 244  
 DB 306 GATATATLSTCIITL-----IVRSCKK 389

RESULT 6  
 AAM94995  
 ID AAM94995 standard; Protein: 431 AA.  
 XX  
 AC AAM94995;  
 XX  
 DT 12-MAY-1999 (first entry)  
 XX  
 DE SAF-2 polypeptide.  
 XX

KM Staladhesin (SAF-2): cellular interaction molecule; cancer; inflammation;  
 KM autoimmunity; allergy; asthma; Rheumatoid arthritis; multiple sclerosis;  
 KM cerebellar degeneration; Alzheimer's disease; Parkinson's disease; AIDS;  
 KM amyotrophic lateral sclerosis; head injury; neurological abnormality;  
 KM septic shock; stroke; osteoporosis; osteoarthritis; ischemia reperfusion;  
 KM cardiovascular disease; kidney; myocardial; hypotension; hypertension;  
 KM myelodysplastic; anemia; infection.  
 OS Homo sapiens.  
 PN P02245/910-A.  
 XX 02 OCT-1998.  
 XX 26 MAR-1998: 98CA-2226910.  
 XX 02-APR-1997: 97US-0041806.  
 XX (SMK) SMITHKLINE BEECHAM CORP.  
 XX Erickson Miller Co., Kofly KK.  
 XX WPI: 1999-106564/10  
 XX N PDB: AAX17772.  
 PS New staladhesin (SAF-2) polypeptide and polynucleotide - useful as  
 PT diagnostic reagents and for prevention and treatment of cancer.  
 XX AIDS, and Alzheimer's disease  
 PS Claim 1: Page 23: 38PP: English.  
 XX This represents a staladhesin polypeptide (SAF-2) which is utilised as  
 CC a cellular interaction molecule. SAF-2 polynucleotides and polypeptides  
 CC are useful for diagnosing susceptibility to diseases and for screening  
 CC for antagonists, agonists. They can also be used to treat diseases  
 CC associated with a SAF-2 imbalance, including cancer, inflammation,  
 CC autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,  
 CC cerebellar degeneration, Alzheimer's disease, Parkinson's disease,  
 CC multiple sclerosis, amyotrophic lateral sclerosis, head injury damage  
 CC and other neurological abnormalities, septic shock, stroke,  
 CC osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,  
 CC cardiovascular diseases, myocardial infarction, AIDS, hypotension,  
 CC hypotension, myelodysplastic syndromes and other abnormalities; aplastic  
 CC anaemia, male pattern baldness, and bacterial, fungal, protozoan and  
 CC viral infections. SAF-2 polypeptides can be used as vaccines and in gene  
 CC therapy to prevent the diseases. SAF-2 polypeptides are useful for  
 CC mapping genes to chromosomes, allowing gene inheritance to be studied  
 CC through linkage analysis.  
 XX Sequence 431 AA:  
 SS Query Match 13.1M; Score 169.5; DB 20; Length 431;  
 SS Best local similarity 23.68, Pval. No. 1.8e 06,  
 SS Matches: 61, Conserved: 53, Mismatches: 92, Indels: 59, Gaps: 11;  
 QY 16 KEMSSVITVNTKNTNTYIIT-TPNSQASITCAVQWHTREELMTREGRDLS--GN 72  
 QY 144 Kqslvstvtatlttptdlitltlshstststltvsvp .....wackstltptmstwsda 134  
 QY 73 KINS-----SSVSSISENNGISFGRD---GRHVSVSVALNTEPP----- 115  
 QY 195 svsspgtltarsvltltkpgdgtstltcqvltptqytlstltstvtldvsypentlmty 254  
 QY 116 -----LISNDQVTEGSSNVKLVNVAANPQAMMMYKNSLLDLEKSHQIQOT 166  
 QY 255 tttatstalttsgstststgtstlttcaavnsnptltstwtstgtstlttststst 307  
 QY 167 SSSVSLSTVVEKFKDNTSTAKSSKLTSTLFFHLITVKKVGV--PTEPIAAV--- 221  
 QY 308 snpplletprlvtrdegetctcraagqststststststststststststststst 367  
 QY 222 VETLTGCTGLAKKKIMLEMK 244

DB 368 dataataltststt1: 1ststck 389  
 RESULT 7  
 ID AAV98565 standard; protein; 848 AA.  
 XX AAV98565;  
 AC AAV98565;  
 XX 07 AUG 2000 (first entry)  
 DE Human NCAM 140KD isoform precursor amino acid sequence.  
 XX NCAM: neural cell adhesion molecule; Ig1: immunoglobulin domain 1;  
 KM neurite outgrowth promoter; proliferation; nerve damage; sclerosis;  
 KM impaired myelination; stroke; Parkinson's disease; memory; schizophrenia;  
 KM Alzheimer's disease; diabetes mellitus; celiac disease; nephrosis;  
 KM treatment; prosthetic nerve guide; treatment; nervous system.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 PN W02001M001-A2.  
 XX 06-APR-2000.  
 XX 4-SEP-1999: 99WR-060500.  
 XX 29-SEP-1998: 98WR-0001232.  
 XX 29 APR 1999: 99DR 0000592.  
 XX (RONN/) RONN T C B.  
 PA (WOLF) FOLF E.  
 PA (HOLM) HOLM A.  
 PA (OLSE) OLSEN M.  
 PA (OSTE) OSTERGAARD S.  
 PA (JENSEN) JENSEN P H.  
 PA (FOUL) FOUJEN F M.  
 PA (SOPKA) SOPKA V.  
 PA (BALB) BALBIS J.  
 PA (BERN) BERLIN V.  
 XX Ronn TCB, Bock F, Holm A, Olsson M, Ostergaard S, Jonson PH,  
 PI Foulsten FM, Sopoka V, Balbais J, Berlin V;  
 XX WFI, 2000 233117/25.  
 XX Compositions that bind neural cell adhesion molecules useful for  
 PT treating disorders of the nervous system and muscles e.g. Alzheimer's  
 PT and Parkinson's diseases -  
 XX Disclosure: Fig 17: 119pp: English.  
 PS This sequence represents the human neural cell adhesion molecule (NCAM)  
 CC amino acid sequence. NCAM is found in three forms, two of which are  
 CC transmembrane forms, while the third is attached via a lipid anchor to  
 CC the cell membrane. All three NCAM forms have an extracellular structure  
 CC consisting of five immunoglobulin domains (I9 domains). The I9 domains are  
 CC numbered 1 to 5 from the N terminal. The invention relates to a compound  
 CC containing a peptide which binds to the NCAM I9 domain. The compound  
 CC binds to NCAM-I9/I92 domains, and is capable of stimulating or promoting  
 CC neurite outgrowth from NCAM presenting cells, and is also capable of  
 CC promoting the proliferation of NCAM presenting cells. The compound may be  
 CC used in the treatment of normal, degenerated or damaged NCAM presenting  
 CC cells. The compound may in particular be used to treat diseases of the  
 CC central and peripheral nervous systems such as post operative nerve  
 CC damage, traumatic nerve damage, impaired myelination of nerve fibres,  
 CC conditions resulting from a stroke, Parkinson's disease, Alzheimer's  
 CC disease, dementia, sepsis, nerve degeneration associated with  
 CC diabetes mellitus, disorders affecting the circadian clock or  
 CC neuro-muscular transmission and schizophrenia. Conditions affecting the  
 CC muscles may also be treated with the compound, such as conditions  
 CC associated with impaired function of neuromuscular connections

CC (e.g. genetic or traumatic shock or traumatic atrophic muscle disorders).  
 CC Conditions of the gonads, pancreas (e.g. diabetes mellitus types I and  
 CC II), kidney (e.g. nephrosis), heart, liver and bowel may also be treated  
 CC using the compound. The compound is used in a prosthetic nerve guide, and  
 CC also to stimulate the ability to learn, and to stimulate the memory of a  
 CC subject.

XX Sequence 848 AA:

Query Match 12.8%; Score 166; DB 21; Length 848;

Best Local Similarity 28.0%; Freq. No. 3e 06; Mismatches 80; Indels 34; Gaps 9;

Matches 58; Conservative 35; Mismatches 80; Indels 34; Gaps 9;  
 QY 41 QASLIC-----AVQNTPEPFIWYREGFVLDKSNKINSSVCVSSISE 86  
 DB 324 qvllceagdpisltwtrstniseektl----dghmvvshary--ssltksiqy 377

QY 87 NDNGISFTC-----RLGRDSVSVSVALNTFPPLLSGNDPQVTEGSGNVKLVCKANKANPO 142  
 DB 378 tdaq-eyictasntlqgd---sqsmylevqyapkigpavayvtwegnqgnltcevfayps 433

QY 143 AQMMWKNSLLDLEK-SHQULQUTSESHULSTIKVEKPDNCTVSCIAKSSLUKTEFLDHF 401  
 DB 434 atlswttrdqqllpsnsmnikikyntrpsaaylevrtpdsendfgcynclavniqgeslefi 493

QY 202 LIAKDKTVGNP-----LEPIAACVVF 224  
 DB 494 lvgad-tpsspsidgyepystaqvfi 519

#### RESULT 8

ID AAB61142 standard; Protein; 404 AA.

AC AAB61142;

DT 30 MAR 2001 (first entry)

DE Human NOV12 protein.

XX Human: NOVX; antiinflammatory; cytosolic; neuroprotective;

KM cerebroprotective; immunomodulator; vulnery; vasotropic; gene therapy;

KM hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;

KM diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.

OS Homo sapiens.

PN WO200075321-A2.

PD 14-DEC-2000.

PE 01-JUN-2000; 2000WU-US15303.

PR 03-JUN-1999; 99US-0137322.

PR 16-MAR-2000; 2000US-0189810.

PR 23-MAR-2000; 2000US-0191158.

PR 30-MAR-2000; 2000US-0193086.

PR 31-MAY-2000; 2000US-0137322.

PA (CORA-) CORAGEN CORP.

PI Shimkets RA, Fernandes E, Herrman J, Vernet C;

DR WPI: 2001-102403/11.  
 DR N-PSDB: AAF27860.  
 XX New N-WX polypeptides and polynucleotides, useful in gene therapy, as a  
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug  
 PT target for treating immune, proliferative and metabolic diseases and  
 PT wound healing  
 PS Claim 1; Page 44-46; 194pp; English.

XX The present sequence is a new isolated polypeptide (NOVX). The NOVX  
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for  
 CC treating or preventing NOVX-associated disorders. They are also useful  
 CC for determining the presence of or a predisposition to a disease  
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.  
 CC These NOVX-associated disorders include hyperplasias, tumours,  
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,  
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral  
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides  
 CC are especially useful in gene therapy. Specifically, NOVX is useful as  
 CC a diagnostic marker or prognostic marker, protein therapeutic and  
 CC antibody target or small molecule drug target to treat disorders in the  
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic  
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive  
 CC disorders), proliferative diseases, or tissue regeneration and  
 CC development (e.g. wound healing or treatment of burns).

SQ Sequence 404 AA:

Query Match 12.5%; Score 162; DB 22; Length 404;

Best Local Similarity 22.4%; Freq. No. 7.e-06; Mismatches 112; Indels 62; Gaps 13;

Matches 64; Conservative 48; Mismatches 112; Indels 62; Gaps 13;  
 QY 14 LPREMTSSVLTVNG-----KTENYILDTTPGSAQLICAVONHTREELMYREGHY-D 67  
 DB 118 mpyktskayltvlygypkqisgisspymegdlmqllctksqspadrtwfkndekld 177

QY 68 LK--SGNKNINSSVCVSS-----ISENDGISTFRLGRDQ-----SVS 104  
 DB 178 vkykdeadanrktflvssslldfrvdsdgvavltcvdhesinatpvyamylethltps 237

QY 105 VSVVLNVTFPEPLLSGNPFYTEGSGNVKLVCKANKANPQAQ--MMWYKN--SSLLDEKSRHQ 162  
 DB 238 vkliptstpt qeqqpllltceskqkplpepvlwkdqgelpdpdr----- 282

QY 163 LVQISEKSLSLIKVEKKNQNIYSCIASSSLKLESLHLVYKKTV-----GVTEPIIA 218  
 DB 283 - mvvsgrdnllflknkngtgrceatnlfgssaeyvllvhdpnaiaagndpdaiaq 340

QY 219 ACVVFETLC -FGLIARRKIKMKMKDKD PSEFAL 254  
 DB 341 givavvfvtlcsifllgrylarthktyltneaksgaedapadadai 386

#### RESULT 9

ID ABB71337 standard; Protein; 1079 AA.

AC ABB71337;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 40803.

KM Drosophila developmental biology, cell signalling, insecticide;

KM pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PE 23 MAR 2001, 2001WO US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PWD, Myers EM;





KM	beta-amyloid peptide; beta-A $\beta$ ; Alzheimer's disease; Downs syndrome;
KX	HCHMA-D; Swedish mutation.
OS	Homo sapiens.
FH	Key
FT	Peptide
FT	Protein
FT	Peptide
FT	Peptide
FT	Peptide
FT	Peptide
FT	Peptide
FT	Domain
PN	US6221645-B1.
XX	24-APR-2001.
PD	
XX	07-JUN-1996; 960S-0660531.
XX	07-JUN-1995; 950S-0480496.
XX	(ELAN-) ELAN PHARM INC.
PA	
XI	Chrystler SMS, Simha S, Keim PS, Anderson JP, Tan H, McConlogue LT;
PI	
XX	WP1: 2001-315578/33.
DR	N-PDB: AAC85809.
XX	
PT	Novel antibody that specifically binds native beta-secretase protein,
PT	useful for raising anti-idiotypic antibodies and for detecting or
PT	diagnosing pathological conditions related to presence of respective
PT	antigens -
XX	
XX	Example: Fig 1. 42pp: English.
CC	This sequence represents beta-secretase protein which is characterized
CC	by an ability to cleave the 695-amino acid iso-type of beta-amyloid
CC	precursor protein (APP) between amino acids 596 and 597. Beta-
CC	secretase is thought to be responsible for the pathogenic processing
CC	of APP to form beta-amyloid peptide (beta-A $\beta$ ) in beta-A $\beta$ related
CC	conditions, e.g., Alzheimer's disease, Downs syndrome, HCHMA-D etc.
CC	Beta-secretase has a molecular weight of 260-300 kD and will bind
CC	to wheat germ agglutinin but not to concanavalin A. Beta-secretase
CC	will cleave both the wild type and the Swedish mutation of APP.
XX	
SQ	Sequence 444 AA:
Dc	Query Match 12.2%; Score 158; DB 23; Length 444;
Dc	Best Local Similarity 20.6%; Pred No. 1.9e-05;
Dc	Matches 67; Conservative 48; Mismatches 109; Indels 102; Gaps 13.
Dc	14 LPREMTSVLTVNGKTE-----NYLDTTPGSQAQLICAVONTHREELLWPEFNPV-D-67
Dc	: : :           : : :       : : :       : : :       : : :
Dc	118 mpyktskayltvlgpekpqisgfsapwmgdlmqtlctsgskpaadlrfrndkeikd 177
Dc	: : :           : : :       : : :       : : :       : : :
Dc	68 LK--SGKKINSSVCSS-----ISENDGISFTGRLGRDQ-----SVS 104
Dc	: : :           : : :       : : :       : : :       : : :
Dc	178 vkylbeedarfkftvssldfrvdstdgyavavcrvheshlnatfpyamqvlelhyps 237
Dc	: : :           : : :       : : :       : : :       : : :
Dc	105 VSVVLNTEFPPLLSGNDPOTVEGSGNVKLVCNVKANPOAO-MMMYKN-SLLDEKSRHO 162
Dc	: : :           : : :       : : :       : : :       : : :
Dc	238 vkiifstgf-----qyghlilticskykplpevlvtukbgjelppqdr-----282
Dc	: : :           : : :       : : :       : : :       : : :
Dc	163 IQQTSEFGJSTIKVKPKPNCTGYCSLAASSLKTESLDHLIVKD-----206
Dc	: : :           : : :       : : :       : : :       : : :

```

Db      293 --mwvsgrelfllfnkcdngcyfreaentfpgssaeyvliwvdpntllpflpsltt 340
               ||:|:| | ||||| | :|:| | |||||
QY      207 ----KTVGVPIEPIAA-----CWVIFLTC---FGLI 232
               ||:|:| | |
Db      341 arvrivvaltrpsrarsrrpnaiaagrgpghaigjvaavvtr lcsfllgry 400
               ||:|:| | | :|:| |
QY      233 ARRKIMKLQMKDKD----PHSETAL 254
               ||:|:| | | :|:| |
Db      401 lathkqyltneakgaedapdadta 426

RESULT 12
AAB50692
ID      AAB50692 standard; Protein; 1415 AA.
AC      AAB50692;
AS      AAB50692;
DE      19-MAR-2001 (first entry)
XX
XX
XX      C. elegans UNC-40 protein dbj ID NO:93.
XX
XX
XX      Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;
XX      protein-protein interaction; identification.
XX
XX      Caenorhabditis elegans.
XX
XX      W0200073420-A2.
XX
XX      07-DEC-2000.
XX
XX      0-2-JUN-2000; 2000W0-EP05108.
XX      1-F-
XX      01-JUN-1999; 99GB-0012755.
XX
XX      (DEVG-) DEVGEN NV.
XX
XX      Van Criekinge W, Roelens I, Bogaert T, Verwaerde P;
XX
XX      WPI: 2001-016508/02.
XX      DR
XX      N-PSDB: AAC90960.
XX
XX      Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
XX      PT a human unc-5H1 cDNA, useful in yeast two hybrid experiments for
XX      PT identifying unknown human cDNAs which encode proteins that interact
XX      PT with the human unc-5C protein -
XX
XX      PS
XX      PS
XX      Disclosure: Page 233-238; 246pp; English.
XX
XX      The present invention describes 3 variants of human unc-5C cDNAs
XX      CC (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
XX      CC unc-5C transcripts, and a human unc-5H1 cDNA which shares homology with
XX      CC the Rattus norvegicus unc-5H1 cDNA. Also described are assays based on
XX      CC protein-protein-interactions between the unc-5 protein and a variety of
XX      CC different interacting proteins. The unc-5C variant cDNAs and unc-5H1
XX      CC cDNA are useful in methods for identifying compounds which reduce or
XX      CC inhibit the ichth phenotype associated with the expression of the
XX      CC unc-5 death domain in yeast. They are also useful in yeast two hybrid
XX      CC experiments for identifying unknown human cDNAs which encode proteins
XX      CC that interact with the human unc-5C protein. AAC90914 to AAC90971 and
XX      CC AAB50646 to AAB50693 represent sequences used in the exemplification of
XX      CC the present invention.
XX
XX      Sequence 1415 AA:
XX
XX
XX      Query Match 12.2%; Score 157.5; DB 22; Length 1415;
XX      Best Local Similarity 27.8%; Pred. No. 9,86-05;
XX      Matches 58; Conservative 35; Mismatches 83; Indels 33; Gaps 10.
XX
XX      1b HEMISSVLIINCKI--ENYILDIIF-----GSASLILVQWHTPEFFELIATPEECR 65
XX      | :|:| | | | | | | | | | | | | | | | | | | | | | | | | |
XX      227 rsgatllvltvleevnevlfltpplqyvgwdeffllleclvasllr-pqyrvlksdrg 265

```







GenCure version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 17, 2002, 13:09:55 ; Search time 86.46 Seconds  
(without alignments)  
336.587 Million cell updates/sec

```

Title: US-09-863-823-b
Perfect score: 1336
Sequence: 1 MAMKSSVIMMGKLLVLL.....KKKIMKLKMKLKLPHSEIAL 262

```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 segs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

```
Database : A.Geneseq.032802.*
1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	311	23.3	66	22	AAU018000	Human immunoglobulin
2	173	12.9	359	22	ABB63920	Drosophila melanog
3	171	12.8	1571	19	AAAM4087	Human Down syndrome
4	170	12.7	1910	19	AAAM4086	Human Down syndrome
5	169.5	12.7	431	20	AAW94995	SAF-2 polypeptide
6	169.5	12.7	431	20	AAW94995	SAF-2 polypeptide
7	166	12.4	848	21	AAV88565	Human NCAM 140KD I
8	162	12.1	494	22	AAH61142	Human N-VL protease
9	161	12.1	1079	22	AAH73137	Drosophila melanog
10	158	11.8	444	20	AAV33741	Beta-secretase. H
11	158	11.8	444	22	AAH47251	Beta-secretase. H

[illegible]



CC protein activity. The sequences can be used as research tools for  
 CC receptors or other signal transduction pathway proteins that interact  
 CC with the polypeptides of the invention and can be used to treat, prevent  
 CC or diagnose various types of disorders such as neurological disorders,  
 CC cardiovascular disorders, gastrointestinal disorders, reproductive  
 CC disorders, immune system disorders, renal disorders, muscular disorders,  
 CC pulmonary disorders, proliferative disorders and cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPU  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 66 AA:

Query Match 23 78: Score 311; DB 22: Length 66;  
 Best Local Similarity 100.0%; Pred. No. 9.3e-20;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 MMYNSSLIDLEKSRHQIOOTSEFOLSTIKYKPKPONGTYSCIAKSLKTESIDFLIV 212  
 DB 1 mmwyssstldeksrhnqststfqlstkrkpkpogtysciakslksldfliv 60

QY 213 K 213  
 DB 61 k 61

RESULT 2

ABB63920 standard; Protein: 359 AA.

AC ABB63920;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 18552.

XX Drosophila: developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL08023.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

XX Disclosure: SEQ ID NO 18552; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB57737-AB872072)  
 CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPU  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 359 AA:

Query Match 12.98: Score 173; DB 22: Length 359;  
 Best Local Similarity 27.38; Pred. No. 8e-07;  
 Matches 50; Conservative 43; Mismatches 85; Indels 32; Gaps 11;

QY 3 MKSSVIMQMGKFLVLVIFLPREMTSSVLTYNCKTENYIDTTPGSOASLCAVQHTRE 62  
 DB 2 wrpsincvswtlllaifvqclaqtrptcstytgqkidi--qgtvefdesvq yake 57

QY 63 EFLIWP-EPRVNDIKSCNKI-----NSSS--VWSSISFNTNIGSTPR-- 104  
 DB 58 yvnlfktdspvflstgctvskdrfsrltydpnsstytqtkldqetdag tytcqv 116

QY 105 LGRQSVSVSVLWVTPPRLSLGNDPOTV--EEGSNKLVCNKKAPQANMMYK--NSSL 161  
 DB 117 lsvtkvsaevklsvrpfvlsdnstgsvasegveqmecyasgyptpltlwrenna 176

QY 162 IDLEKSRHQIOOTSEFOLSTIKYKPKPONGTYSCIAKSL 201  
 DB 177 lptdsa-----tyvgnrlrlvsvtkedrgtgyvvaingv 210

RESULT 3

AAW42087 standard; Protein: 1571 AA.

AC AAW42087;

DT 28-SEP-1998 (first entry)

XX Human Down syndrome-cell adhesion molecule DS CAM2.

XX DS-CAM2: Down syndrome-cell adhesion molecule; neural cell;

XX signal transduction; trisomy 21; mental retardation;

XX holoprosencephaly; corpus callosum agenesis;

XX schizencephaly; diagnosis; assay; human.

OS Homo sapiens.

PN WO9817795-A1.

PD 30-APR-1998.

PF 23-OCT-1997; 97WO-US19547.

PR 25-OCT-1996; 96US-0029322.

PA (CEDA-) CEDARS SINAI MEDICAL CENT.

PI Korenberg JR;

DR WPI; 1998-271791/24.

DR N-PSDB; AAW31988.

XX New isolated Down's Syndrome-cell adhesion molecule - used to  
 PT develop products for detection, diagnosis and therapy of  
 PT developmental and neurological abnormalities

XX Claim 2; Page 90-95; 109pp; English.

XX This polypeptide comprises Down syndrome-cell adhesion molecule  
 CC DS-CAM2, an extracellular soluble protein belonging to a novel  
 CC subclass of the Ig superfamily with highest homology to neural cell  
 CC adhesion molecules. Its amino acid sequence was deduced from cDNA  
 CC clones (see AAW31982) isolated from a trisomy 21 fetal brain library.  
 CC It is a splice variant of membrane-bound DS-CAM1 (see AAW42086), and  
 CC lacks the entire transmembrane domain of DS-CAM1. The invention  
 CC provides human and murine DS-CAM nucleic acid sequences (see also

Sequence 1571 AA;

107

[illegible]



PT New isolated Down's Syndrome-cell adhesion molecule - used to  
 PT develop products for detection, diagnosis and therapy of  
 PT developmental and neurological abnormalities  
 PS Claim 2; Page 73-78; 109pp; English.  
 XX  
 CC This polypeptide comprises Down syndrome-cell adhesion molecule  
 CC DS-CAM1, a cell surface glycoprotein belonging to a novel subclass  
 CC of the Ig superfamily with highest homology to neural cell adhesion  
 CC molecules. Its amino acid sequence was deduced from cDNA clones  
 CC (see AAV11981) isolated from a trisomy 21 foetal brain library. A  
 CC splice variant, DS-CAM2 (see AAM42087), which is non-membrane bound  
 CC was also identified. The invention also provides human and murine  
 CC DS-CAM nucleic acid sequences (see also AAV11985-88), expression  
 CC vectors and host cells, transgenic animals, antibodies, antisense  
 CC oligonucleotides, and primers derived from DS-CAM nucleic acid.  
 CC DS-CAM polypeptides are associated with developmental and  
 CC neurological processes. They can be used in e.g. neural prosthetic  
 CC devices used in entubulation methods of repairing (regenerating)  
 CC damaged or severed peripheral nerves, and also in bioassays to  
 CC identify agonists and antagonists. The products can also be  
 CC used in detection, diagnosis and therapy of developmental and  
 CC neurological abnormalities such as Down syndrome, mental  
 CC retardation, holoprosencephaly, agenesis of the corpus callosum,  
 CC or schizencephaly.  
 CC  
 XX  
 SQ Sequence 1910 AA.  
 Query Match 12.7%; Score 170; DB 19; Length 1910;  
 Best Local Similarity 29.8%; Pred. No. 1.3e 05;  
 Matches 57; Conservative 31; Mismatches 69; Indels 34; Gaps 10;  
 QY 23 PREMTSSVLVYNGKTEVYIIDTTPGASALICAVQNHTRREELWYREGRYDLKS--GN 80  
 DB 320 PRKXKSSV-----GASQVSLSCSVG--TEDQELSWYR--YQEL--LPPYKXV 360  
 QY 82 -----INSSVCVSSISENDNGISFTGRLGRQSVS---VSVLVNVPFLLSQNDPQVE 134  
 DB 361 RITGIMHNLIMDMVKSQGG--AYQCTVTRKDKLSAQDYQVVEDGTPRLISAFSKVYS 419  
 QY 135 EGSNWKLVQCNKANPQAKQMMWYKNSLLDEKSRHQIQQ--TSES---POLSTTYEKPD 189  
 DB 420 PAEPVSLMCMVKGPIPTLTWLTDDPI--LKQGSCHRISQMLTSEGIVVSYLINSSEQVD 478  
 QY 190 NGTYSCTAKSS 200  
 DB 479 GGVYRCANNS 489  
 RESULT 5  
 AAM94996  
 ID AAM94996 standard; Protein; 431 AA.  
 AC AAM94996;  
 DT 12-MAY-1999 (first entry)  
 DE SAF-2 polypeptide (derived from EST).  
 XX Staloathesis; SAF-2; cellular interaction molecule; cancer; inflammation;  
 XX autoimmunity; allergy; asthma; rheumatoid arthritis; multiple sclerosis;  
 XX cerebellar degeneration; Alzheimer's disease; Parkinson's disease; AIDS;  
 XX amyotrophic lateral sclerosis; head injury; neurological abnormality;  
 XX septic shock; stroke; osteoporosis; osteoarthritis; ischemia reperfusion;  
 XX cardiovascular disease; kidney; myocardial; hypotension; hypertension;  
 XX myelodysplastic; anaemia; infection.  
 XX Homo sapiens.  
 XX OS  
 XX CA226910-A.  
 XX PD 02-OCT-1998.

XX  
 FF 26-MAR-1998; 98CA 226910.  
 XX  
 PR 02-APR-1997; 97US-004186.  
 XX  
 PA (SMK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Erickson-Miller CL, Kikly KK;  
 XX  
 DR WPI; 1999-106564/10.  
 XX  
 PT New staloathesis (SAF-2) polypeptide and polynucleotide - useful as  
 PT diagnostic reagents and for prevention and treatment of cancer,  
 PT AIDS, and Alzheimer's disease  
 PS Claim 14; Page 23-24; 38pp; English.  
 XX  
 CC The invention relates to a staloathesis polypeptide (SAF-2) which is  
 CC utilised as a cellular interaction molecule. SAF-2 polynucleotides and  
 CC polypeptides are useful for diagnosing susceptibility to diseases and for  
 CC screening for antagonists, agonists. They can also be used to treat  
 CC diseases associated with a SAF-2 imbalance, including inflammation,  
 CC autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,  
 CC cerebellar degeneration, Alzheimer's disease, Parkinson's disease,  
 CC multiple sclerosis, amyotrophic lateral sclerosis, head injury damage  
 CC and other neurological abnormalities, septic shock, cancer, stroke,  
 CC osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,  
 CC cardiovascular diseases, myocardial infarction, AIDS, hypotension,  
 CC hypertension, myelodysplastic syndromes and other abnormalities, aplastic  
 CC anaemia, male pattern baldness, and bacterial, fungal, protozoan and  
 CC viral infections. SAF-2 polypeptides can be used as vaccines and in gene  
 CC therapy to prevent the diseases. SAF-2 polypeptides are useful for  
 CC mapping genes to chromosomes, allowing gene inheritance to be studied  
 CC through linkage analysis.  
 CC  
 XX  
 SQ Sequence 431 AA;  
 Query Match 12.7%; Score 169.5; DB 20; Length 431;  
 Best Local Similarity 23.6%; Pred. No. 2e-06; Mismatches 92; Indels 59; Gaps 11;  
 Matches 63; Conservative 53;  
 QY 24 REMTSSVLVYNGKTEVYIIDTTPGASALICAVQNHTRREELWYREGRYDLKS--GN 80  
 DB 144 KQLSVFTALTRPDLLIIGLLESQHSRLTCSVP-----WACKQGTPTMNSWIGA 194  
 QY 81 KINS-----SSVCVSSISENDNGISFTGRL--GRDSVSVSVLVNVPFLLSQNDPQVE 123  
 DB 195 SVSSPGPTARSSVLTLPKPYGHSILTCQVLTGVTLSVTRLDVSPYPMWLTMTV 254  
 QY 124 -----LTSGNDPQVEGSNVKNLVQCNKANPQAKQMMWYKNSLLDEKSRHQIQQ 174  
 DB 255 FQGDATALGNGSSLSVLEQSLITVCVNSPPARLSWIGS--LILGSR-----S 307  
 QY 175 SESFOLSTTYEKPNDCTYSIAKSSILKTESLDFHLVMDKTVGV--PIEPIIACV--- 229  
 DB 308 SNPGLLLEPRVTRDEGETICRQAGNSQHSLSLSLQNGQSLRSPQSVYLAAVQGA 367  
 QY 230 -----VIFLTLCFGLIARRKKIMKLCMK 252  
 DB 368 GATALATSLTCLIF-----IVRSCK 389  
 RESULT 6  
 AAM94995  
 ID AAM94995 standard; Protein; 431 AA.  
 AC AAM94995;  
 DT 12-MAY-1999 (first entry)  
 DE SAF-2 polypeptide.  
 XX  
 XX

KW Stalodhesin; SAF-2; cellular interaction molecule; cancer; inflammation;  
 KW autoimmunity; allergy; asthma; rheumatoid arthritis; multiple sclerosis;  
 KW cerebellar degeneration; Alzheimer's disease; Parkinson's disease; AIDS;  
 KW amyotrophic lateral sclerosis; head injury; neurological abnormality;  
 KW septic shock; stroke; osteoporosis; osteoarthritis; ischemia reperfusion;  
 KW cardiovascular disease; kidney; myocardial infarction; hypertension;  
 KW myelodysplastic anemia; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN CA226910-A.  
 XX  
 PD 02-OCT-1998.  
 XX  
 XX  
 PE 26-MAR-1999; 96CA-226910.  
 XX  
 PR 02-APR-1997; 97OS-004180C.  
 XX  
 FA (SMF) SPITHELINE; BERNHAM (9-97).  
 XX  
 PI Erickson-Miller CL, Kikly KK.  
 XX  
 DR WP1: 1999-106564/10.  
 DR N-PSDB; AAX17772.  
 XX  
 PT New Stalodhesin (SAF-2) polypeptide and polynucleotide useful as  
 PT diagnostic reagents and for prevention and treatment of cancer,  
 PT AIDS, and Alzheimer's disease.  
 XX  
 PS Claim 1; Page 23; 48pp; English.  
 XX  
 CC This represents a stalodhesin polypeptide (SAF-2) which is utilized as  
 CC a cellular interaction molecule. SAF-2 polynucleotides and polypeptides  
 CC are useful for diagnosing susceptibility to diseases and for screening  
 CC for antigens, antibodies. They can also be used to treat diseases  
 CC associated with a SAF-2 imbalance, including cancer, inflammation,  
 CC autoimmunity, allergy, asthma, rheumatoid arthritis, CNS inflammation,  
 CC cerebellar degeneration, Alzheimer's disease, Parkinson's disease,  
 CC multiple sclerosis, amyotrophic lateral sclerosis, head injury damage  
 CC and other neurological abnormalities, septic shock, stroke, stroke,  
 CC osteoporosis, osteoarthritis, ischemia reperfusion injury, kidney, liver,  
 CC cardiovascular diseases, myocardial infarction, AIDS, hypertension,  
 CC hyperkalemia, myelodysplastic syndromes and other abnormalities, aplastic  
 CC anaemia, male pattern baldness, and bacterial, fungal, protozoan and  
 CC viral infections. SAF-2 polypeptides can be used as vaccines and in gene  
 CC therapy to prevent the diseases. SAF-2 polypeptides are useful for  
 CC mapping genes to chromosomes, allowing gene inheritance to be studied  
 CC through linkage analysis.  
 CC  
 XX Sequence 41; AA:  
 SO  
 Query Match 127W; Score 169.5; DB 20; Length 431;  
 Host Local Similarity 23.6%; Fred. No. 20 95;  
 Matches 63; Conservative 53; Mismatches 32; Indels 59; Gaps 11,  
 UY 24 KEMTSVLTAVNKRTNTNLTIT-THSSQSLICAVONTREBELLYTEDEGVILKS--GN 80  
 DB 144 kqfscvltalhpdlittpgtresstnltitpfp --- waakqnfepmlswlga 194  
 UY 81 KINS SSVGVSSISENENISFTGRD--GRDSVSVSVNLTTPF----- 124  
 DB 195 wssghpplttusa::ltpkphdhlstlstepllpqvtltstclldvsfpwmlnly 254  
 UY 124 -----LSSNPEVTPVPSNVKLVWVFAVWQAGMMYKNSLLDEKSPHQIQT 174  
 DB 255 lqdlstslstlndst 307  
 UY 175 STSPYNTKVKKPTNTVSTAKSSIKTSILRLVKECTVY -PIETINACY--- 229  
 DB 308 lqdlstslstlndst 367  
 UY 230 VETTLTFTGGLAKRKKIMKLMKR 252

DB 368 data:distc111 -LWSTCR 389  
 RESULT 7  
 ID AAY88565 standard; protein: 848 AA.  
 XX  
 AC AAY88565;  
 XX  
 DE 07-AUG-2000 (first entry)  
 XX  
 DE Human NCAM 140kd isoform precursor amino acid sequence.  
 XX  
 KW NCAM; neural cell adhesion molecule; Ig1; immunoglobulin domain 1;  
 KW neurite outgrowth promoter; proliferation; nerve damage; sclerosis;  
 KW impaired cognition; stroke; Parkinson's disease; memory; schizophrenia;  
 KW Alzheimer's disease; diabetes mellitus; circulation; kidney nephrosis;  
 KW treatment; prosthetic nerve tides; treatment; nervous system.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W020001801 A2.  
 XX  
 PD 06-APR-2000.  
 XX  
 PE 23-SEP-1999; 99W0-DK00500.  
 XX  
 PR 29-SEP-1998; 98UK-0001212.  
 PR 29-APR-1999; 99DK-0000592.  
 XX  
 PA (PNNN) RONN L C B.  
 PA (BOCK) BOCK E.  
 PA (HOLM) HOLM A.  
 PA (OLSE) OLSEN M.  
 PA (OSTE) OSTERGAARD S.  
 PA (JENSE) JENSEN P H.  
 PA (POUL) POULSEN F M.  
 PA (SORO) SOROKA V.  
 PA (RALE) RALETS I.  
 PA (BERE) BEREZIN V.  
 XX  
 PI Ronn LCH, Bock E, Holm A, Olsen M, Ostergaard S, Jensen PH;  
 PI Poulsen FM, Soroka V, Ralets I, Berezin V;  
 XX  
 PS WPI: 2000-293111/25.  
 XX  
 CC compositions that bind neural cell adhesion molecules useful for  
 CC treating disorders of the nervous system and muscles e.g. Alzheimer's  
 CC and Parkinson's diseases.  
 XX  
 PS Disclosure: Fig 17; 119pp; English.  
 XX  
 CC This sequence represents the human neural cell adhesion molecule (NCAM)  
 CC amino acid sequence. NCAM is found in three forms, two of which are  
 CC transmembrane forms, while the third is attached via a lipid anchor to  
 CC the cell membrane. All three NCAM forms have an extracellular structure  
 CC consisting of five immunoglobulin domains (Ig domains). The Ig domains are  
 CC numbered 1 to 5 from the N-terminal. The invention relates to a compound  
 CC containing a peptide which binds to the NCAM Ig1 domain. The compound  
 CC binds to NCAM Ig1/192 domains, and is capable of stimulating or promoting  
 CC neurite outgrowth from NCAM presenting cells, and is also capable of  
 CC promoting the proliferation of NCAM presenting cells. The compound may be  
 CC used in the treatment of normal, degenerated or damaged NCAM presenting  
 CC cells. The compound may in particular be used to treat diseases of the  
 CC central and peripheral nervous systems such as post operative nerve  
 CC damage, traumatic nerve damage, impaired myelination of nerve fibres,  
 CC conditions resulting from a stroke, Parkinson's disease, Alzheimer's  
 CC disease, dementia, sclerosis, nerve degeneration associated with  
 CC diabetes mellitus, disorders affecting the circadian clock or  
 CC neuro muscular transmission and schizophrenia. Conditions affecting the  
 CC muscles may also be treated with the compound, such as conditions  
 CC associated with impaired function of neuromuscular connections

CC (e.g. genetic or traumatic shock or traumatic atrophic muscle disorders).  
 CC Conditions of the gonads, pancreas (e.g. diabetes mellitus types I and  
 CC II), kidney (e.g. nephrosis), heart, liver and bowel may also be treated  
 CC using the compound. The compound is used in a prosthetic nerve guide, and  
 CC also to stimulate the ability to learn, and to stimulate the memory of a  
 CC subject.  
 XX  
 XX Sequence 848 AA.

Query Match 12.4%; Score 166; DB 21; Length 848;  
 Best Local Similarity 28.0%; Pred. No. 1e-05;  
 Matches 58; Conservative 35; Mismatches 80; Indels 34; Gaps 9;

QY 49 QASLIC-----AVQNTREBELLYREGRVLDKSNKSNSSVCVSSISE 94  
 Db 324 qvltceasgdpilswltwtsttrnlsssekl-----dghmvrsharv--slltklsqy 377  
 QY 95 NDNGISFTC-----RLGRDQSVSVLVNTPFPPLLSGNDFOVEGGSNWKVCNKAMPQ 150  
 Db 378 tdag eyictasntiqd-----sqsmylevgyapkigqvayvtywegqnlcevfayps 433  
 QY 151 AUMMMWKNSLLDLEK-SKHUJLVQVTSKSHLSITVEKPKUNSTYSCLAKSLKTESLDLPH 209  
 Db 434 atlswtirtdgqllpsnynskikynpsaaylevtpdsendfgynctaavnigqeslfti 493  
 QY 210 LIYKDKTVGP-----IEPIAACVYIF 232  
 Db 494 lvgad-lpsspsidgyepystaqvft 519

RESULT 8  
 AAB61142  
 ID AAB61142 standard; Protein; 404 AA.  
 AC AAB61142;  
 XX  
 DT 30-MAR-2001 (first entry)  
 XX  
 DE Human NOV12 protein.  
 XX  
 OS Human, NOVX; antiinflammatory; cytosolic; neuroprotective;  
 KM cerebroprotective; immunomodulator; vulnery; vasotropic; gene therapy;  
 KM hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;  
 KM diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200075321-A2.  
 PD 14-DEC-2000.  
 XX  
 PF 01-JUN-2000; 2000WO-0515403.  
 XX  
 PR 03-JUN-1999; 390S-0137322.  
 PR 16-MAR-2000; 2000US-0189810.  
 PR 22-MAR-2000; 0191158.  
 PR 30-MAR-2000; 2000US-0193086.  
 PR 31-MAY-2000; 2000US-0137322.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkets RA, Fernandes E, Herrman J, Vernet C;  
 DR WPI: 2001-102403/11.  
 DR N-PSDB: AAF27860  
 XX  
 XX New NOVX polypeptides and polynucleotides, useful in gene therapy, as a  
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug  
 PT target for treating immune, proliferative and metabolic diseases and  
 PT wound healing  
 XX  
 PS Claim 1; Page 44-46; 194pp; English.

XX  
 CC The present sequence is a new isolated polypeptide (NOVX). The NOVX  
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for  
 CC treating or preventing NOVX associated disorders. They are also useful  
 CC for determining the presence of or a predisposition to a disease  
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.  
 CC These NOVX-associated disorders include hyperplasias, tumours,  
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,  
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral  
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides  
 CC are especially useful in gene therapy. Specifically, NOVX is useful as  
 CC a diagnostic marker or prognostic marker, protein therapeutic and  
 CC antibody target or small molecule drug target to treat disorders in the  
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic  
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive  
 CC disorders), proliferative diseases, or tissue regeneration and  
 CC development (e.g. wound healing or treatment of burns).  
 XX  
 XX Sequence 404 AA.

Query Match 12.1%; Score 162; DB 22; Length 404;  
 Best Local Similarity 22.4%; Pred. No. 4e-06;  
 Matches 64; Conservative 48; Mismatches 112; Indels 62; Gaps 13;

QY 22 LPREMTSSVITVNG-----KTENYILDTPGSOALICAVQNTREBELLYREGRV D 75  
 Db 118 mpvrtvsgyrtvtygrygrypjtsgfsspmegdlmqtkrsgskpaadtvtvndvlelvd 177  
 QY 76 LK--SCNKINSSVCVSS-----ISENDNGISFTCRLGRDQ-----SVS 112  
 Db 178 vkykdeedankrtftvsstldirvdsddqvavivrcvhdeshlnapqamyleihytp 237  
 QY 113 VSVVLNTPPPLLSGNDFOVEGGSNWKLVCKVKNPQAO MMWYKN-SLLDLEKSRHO 170  
 Db 238 vkliptpfp-----qegqpllltccskqkplpepylwtkdqgelppdr----- 282  
 QY 171 IQOTSESFOLSITKVEKPNNGTYSCTJAKSLKTESLDLFLYKDKTV-----GVPEPIIA 226  
 Db 283 --mvvsqrelnllfinktknagfrcacatnigssaaeyllvthdpna3agunqphallg 340  
 QY 227 --ACVVFFLLD-----FCLIAKRRKIMKLMCKDKD-----FNSERIAL 262  
 Db 341 glvaavvfvltcsifllqrylarhbkqyltneakgaedaddat 386

RESULT 9  
 ABB71337  
 ID ABB71337 standard; Protein; 1079 AA.  
 XX  
 AC ABB71337;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 40803.  
 XX  
 KM Drosophila; developmental biology; cell signalling, insecticide;  
 KM pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 21-MAR-2000; 2000US-391337P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;



```

KM beta-amyloid peptide; beta-AP; Alzheimer's disease; Downs syndrome;
KM HCMA-D; Swedish mutation.
XX
XX Homo sapiens
XX
XX Key Location/Qualifiers
XX Peptide 1..23
XX /note= "Putative signal peptide"
XX Protein 24..444
XX /note= "Mature beta-secretase"
XX Peptide 125..141
XX /label= Peptide 1
XX Peptide 190..200
XX /label= Peptide 5
XX Peptide 240..259
XX /label= Peptide 2
XX Peptide 261..272
XX /label= Peptide 3
XX Domain 377..399
XX /note= "Putative transmembrane domain"
XX
XX US6221645-B1.
XX
XX 24-APR-2001.
XX
XX 07-JUN-1996; 96US-0660531.
XX
XX 07-JUN-1995; 95US-0480498.
XX
XX (ELAN-) ELAN PHARM INC.
XX
XX Chrysler SMS, Sinha S, Keim ES, Andersen JP, Fan H. Molecular biology of
XX WPI: 2001-315578/33.
XX N-PSDB: AAC85809.
XX
XX Novel antibody that specifically binds native beta-secretase protein,
XX useful for raising anti-idiotypic antibodies and for detecting or
XX diagnosing pathological conditions related to presence of respective
XX antigens -
XX
XX Example: Fig 1; 42pp; English.
XX
XX This sequence represents beta-secretase protein which is characterized
XX by an ability to cleave the 695-amino acid isoform of beta-amyloid
XX precursor protein (APP) between amino acids 596 and 597. Beta-
XX secretase is thought to be responsible for the pathogenic processing
XX of APP to form beta-amyloid peptide (beta-AP) in beta-AP related
XX conditions, e.g. Alzheimer's disease, Downs syndrome, HCMA-D etc.
XX Beta-secretase has a molecular weight of 260-300 kD and will bind
XX to wheat germ agglutinin but not to concanavalin A. Beta-secretase
XX will cleave both the wild type and the Swedish mutation of APP.
XX
XX Sequence 444 AA:
SQ
Query Match 11.8%; Score 158; DB 22; Length 444;
Best Local Similarity 20.6%; Pred. No. 2,1e-05,
Matches 67; Conservative 48; Mismatches 109; Indels 102; Gaps 13.
OY 22 LPRENTSSVLTVNGKTE-----NYLDTTGSASLCAVQNNRRELLWYKEGV-D 75
DB 118 mpvkskaylvtlvypkqslgfsppmegdlmqllctksgskpadltwfkndkeikd 177
OY 76 LK--SGNKINSSSVSVS-----ISENDNGISPTGRLGRQ-----SVS 112
DB 178 vkylkeedanrkktfvsstldifwdrsdgyavavictvdhesinatpavmqvleinytps 237
OY 113 VSVLVNTPFPPLLSGNDPQVEEGSNVLYCNVKNAPQAO-MMYKXN-SSLDLLEKSRHQ 170
DB 238 vkllpstrfpf-----geqgpllllccesqkplpcepvlwtkdggelppdpr----- 282
OY 171 IOGTSESQSLTKVEKEDNGTYSICIAKSLKTESLDFHLIVK----- 214

```

```

DB 283 --mvvsgremlilflnkrtdnglyrceatnltqgsaeyvlvdpvnpilptiipstc 340
OY 215 ----KVCVPIEPIIAA-----CWFPIITLC-----FCLT 240
DB 341 atvtvtvattspsttsatsttsirdnalagqpdpalligivavvvtlclslilgry 400
OY 241 ARKKIMKLCMKDKD----PHSETAL 262
DB 401 larkgtyllneakgaedapadatal 426
RESULT 12
AAB50692
ID AAB50692 standard; Protein; 1415 AA.
XX
AC AAB50692;
XX
DT 19-MAR-2001 (first entry)
XX
XX C. elegans UNC-40 protein Sfw 1D NO:93.
XX
XX Human; Caenorhabditis elegans; UNC-5; splice variant; nematode worm;
XX protein-protein interaction; identification.
XX
XX Caenorhabditis elegans.
XX
XX W0200073328-A2.
XX
XX 07-DEC-2000.
XX
XX 07-JUN-2000; ZmWU-EP0108.
XX
XX 01-JUN-1999; 99GB-0012755.
XX
XX (DEVCN-) DEVGEN NV.
XX
XX Van Criekeinge W, Roelens I, Bogaert T, Verwaerde P.
XX WPI: 2001-016508/02.
XX N-PSDB: AAC90960.
XX
XX Three variants of human unc-5C cDNAs (unc-5Cb, unc-5Cc and unc-5C8) and
XX a human unc-5HS1 cDNA, useful in yeast two hybrid experiments for
XX identifying unknown human cDNAs which encode proteins that interact
XX with the human unc-5C protein -
XX
XX Disclosure: Page 233-238; 246pp; English.
XX
XX The present invention describes 3 variants of human unc-5C cDNAs
XX (unc-5Cb, unc-5Cc and unc-5C8) which correspond to alternatively spliced
XX unc-5C transcripts, and a human unc-5HS1 cDNA which shares homology with
XX the Rattus norvegicus unc-5HS1 cDNA. Also described are assays based on
XX protein-protein interactions between the unc-5C protein and a variety of
XX different interacting proteins. The unc-5C variant cDNAs and unc-5HS1
XX cDNA are useful in methods for identifying compounds which reduce or
XX inhibit the lethal phenotype associated with the expression of the
XX unc-5C death domain in yeast. They are also useful in yeast two hybrid
XX experiments for identifying unknown human cDNAs which encode proteins
XX that interact with the human unc-5C protein. AAC90914 to AAC90971 and
XX AAB50646 to AAB50693 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 1415 AA:
SQ
Query Match 11.8%; Score 157.5; DB 22; Length 1415;
Best Local Similarity 27.8%; Pred. No. 0.00011;
Matches 58; Conservative 35; Mismatches 83; Indels 33; Gaps 10;
OY 24 REMTSSVLTVNGKTE--ENVYLDTP-----GSOASLCAVQNNRRELLWYREGGR- 73
DB 227 rsgtarltvtetvsnelvtltprlgyvcegdellcevtasllr p3:rwkdsrql 285

```



Query Match 11.6%; Score 155; DB 22; Length 595;  
 Best Local Similarity 23.2%; Pred. No. 5.7e-05;  
 Matches 63; Conservative 46; Mismatches 84; Indels 78; Gaps 12;

39 NYILD-----TPGQAS-----LICAVONHREBELLYREGR----- 73  
 258 NYLDKSLVHVALHPLIFSLIPGLLESHPRILTSVP-----WACEQIPPLIC 308  
 74 ---VDLKGSKKINSSVCSISSE-NDNGISFTCR-GRQSVSVYLVNTPP- 123  
 309 WMAQSVSIPILTRSSMISLIPQDNGTSLCQVLPQAGVTRMTRVRIISYPPQNL 368  
 124 -----LISGNDPEVIEESNVKLVNWKANPQAGMMVKNSSLDLEKSKHQ 170  
 369 TMTVFQDGTASTLTNGSALSVEGSHLVCAVDSNPARTLSWTWS--LTLSPS--- 423  
 171 IQQTSEFQSLITKVEKPDNGTTSCTAKSSLTESLDFLIYKDKTVGV--PIEPI----- 224  
 424 ---GSSNLQVLEIPRVNKKDQGLTCRAQPLGSHISLSIQEYTKMPISQVTLGA 481  
 225 ---IACVVFELTLCFGLIARRKKIMKLCMK 252  
 482 FGAGATAIVFLYFCILFV-----VRSCK 507

RESULT 15

AAG77857  
 ID AAG77857 standard; Protein; 537 AA.

XX AAG77857;

DT 11-JAN-2002 (first entry)

XX Mutant protein mMusK-RI.

XX Mutated muscle specific tyrosine kinase receptor; selectable marker;

KM mMusK-R; human; mMusK-R; antibody; mutant; muten; immunoselection.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..19 /note= "Signal peptide of mMusK-R"

FT Domain 20..493 /note= "Extracellular domain"

FT Domain 494..515 /note= "Transmembrane domain"

XX MO200172834-A1.

PD 04-OCT-2001.

XX 28-MAR-2001; 2001WO-EP03543.

XX 30-MAR-2001; 2001US-0539248.

XX (NOVS ) NOVARTIS AG.

XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX P1PIPIG SD, Veres G;

XX WPI; 2001-626257/72.

PT Identifying genetically modified mammalian cells comprising a nucleic  
 PT acid sequence encoding a mutated muscle-specific tyrosine kinase  
 PT receptor provides a selectable marker for transformation

XX Claim 5; Page: 52pp; English.

XX The sequence represents the mutant protein mMusK-RI. The invention

CC concerns a method of using a muscle specific tyrosine kinase receptor  
 CC molecule (MusK-R) or a mutated MusK-R (mMusK-R) thereof as a selectable  
 CC marker in mammalian cells, particularly human cells. Preferred markers  
 CC are mMusK-R's incapable of signal transduction and preferably molecules  
 CC wherein the intracellular domain has been modified by deletion of the  
 CC signalling region. The invention also relates to a method of identifying  
 CC genetically modified mammalian cells including introducing a MusK-R into  
 CC a target cell as a selectable marker. Further, a method for the  
 CC immunoselection of transduced mammalian cells is disclosed comprising  
 CC identifying the transduced cells by incubation of the cells with an  
 CC antibody which recognises and binds specifically to a MusK-R or mMusK-R  
 CC thereof.  
 CC Note: The present sequence is not shown in the specification but is  
 CC derived from the human wild-type muscle specific tyrosine kinase receptor  
 CC protein shown in Figure 2 (AAG77856).

SO Sequence 537 AA;

Query Match 11.6%; Score 154.5; DB 22; Length 537;  
 Best Local Similarity 25.2%; Pred. No. 5.4e-05;  
 Matches 56; Conservative 41; Mismatches 82; Indels 43; Gaps 11;

QY 15 LLIVYLF-----LPRF--MTSSVLTYNCKTENYILDTPGQASLILCAVONHREBELLY 67  
 DB 12 ILLVATSGREKIPKAPVITIPETVDAIVEV-----ALIMCAVESYPQPEISW 61

QY 68 YREGGRV---DLKSGKINSSVCSISSENDNGISFTCRD-----QSVSVSV 116  
 DB 62 TRNKILIKLIDTYSIRENGIILLISVEDSDGIYCTLANNGYGAUESGALQVYKMK 120

QY 117 LNTFFPLLSGNDPOTVEGSKNKLVCNWKANPQAGMMVKNSSLDLEKSKHQIQQTSE 176  
 DB 121 PKLITRPDI-----NVKILEGLKAVIPCTGMPKPSVSVKGDSPJ-RENRIAV---LE 171

QY 177 SPQSLITKVEKPDNGTTSCTAKSSLTESLDFLIYKDKTVGV--PIEPI----- 224  
 DB 172 SGSLINHVQKEDAGYRCVAKNSIGTAYSKVKLEFEVFAR 213

Search completed: September 17, 2002, 13:09:56  
 Job time: 11323 sec

. . .

---